BAYESIAN MODEL SELECTION USING EXACT AND APPROXIMATED POSTERIOR PROBABILITIES WITH APPLICATIONS TO STAR DATA

A Dissertation

by

SURIANI POKTA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2004

Major Subject: Statistics

BAYESIAN MODEL SELECTION USING EXACT AND APPROXIMATED POSTERIOR PROBABILITIES WITH APPLICATIONS TO STAR DATA

A Dissertation

by

SURIANI POKTA

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

Jeffrey D. Hart
(Chair of Committee)

James A. Calvin
(Member)

Marina Vannucci
(Member)

Douglas A. Hensley
(Member)

Michael T. Longnecker
(Head of Department)

August 2004

Major Subject: Statistics

ABSTRACT

Bayesian Model Selection Using Exact and Approximated
Posterior Probabilities with Applications to Star Data. (August 2004)
Suriani Pokta, B.S., Mathematics, University of California, Los Angeles;
M.A., Statistics, University of California, Berkeley
Chair of Advisory Committee: Dr. Jeffrey D. Hart

This research consists of two parts. The first part examines the posterior probability integrals for a family of linear models which arises from the work of Hart, Koen and Lombard (2003). Applying Laplace's method to these integrals is not entirely straightforward. One of the requirements is to analyze the asymptotic behavior of the information matrices as the sample size tends to infinity. This requires a number of analytic tricks, including viewing our covariance matrices as tending to differential operators. The use of differential operators and their Green's functions can provide a convenient and systematic method to asymptotically invert the covariance matrices. Once we have found the asymptotic behavior of the information matrices, we will see that in most cases BIC provides a reasonable approximation to the log of the posterior probability and Laplace's method gives more terms in the expansion and hence provides a slightly better approximation. In other cases, a number of pathologies will arise. We will see that in one case, BIC does not provide an asymptotically consistent estimate of the posterior probability; however, the more general Laplace's method will provide such an estimate. In another case, we will see that a naive application of Laplace's method will give a misleading answer and Laplace's method must

be adapted to give the correct answer. The second part uses numerical methods to compute the "exact" posterior probabilities and compare them to the approximations arising from BIC and Laplace's method.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to Dr. Jeffrey D. Hart, my advisor, for his kind support and encouragement throughout the years. I would also like to extend my thanks to Drs. Calvin, Chen, Claeskens, Cline, Dahm, Fan, Longnecker, Spiegelman, Vannucci and Wehrly for their generosity and understanding. Lastly, I wish to thank Veera Baladandayuthapani, Karin Chu, Mickey Dunlap, Scott Holan, Jeesun Jung, Gosia Leyk, Daisy Liu, Yolanda Muñoz, Marilyn Randall, Nick Rose and Christie Spinka for their help and friendship.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURE Page

CHAPTER I

INTRODUCTION

In the statistical analysis of data, model selection is often the first step upon which further inferences are based and is therefore a crucial step in many statistical applications. If we regard the likelihood, or more commonly -2 times the log likelihood, evaluated at the Maximum Likelihood Estimators (MLEs) for the model as the sole measure of how well a model fits the data, then a larger model with more parameters will always appear to fit better. This is clearly not appropriate, since models with too many parameters are difficult to interpret, more difficult to implement, and tend to lose touch with reality. The most obvious way of correcting this flaw is to add a term to penalize the inclusion of additional parameters. While a large number of penalty terms could be envisioned, there are several penalty terms which have interesting motivations.

The first such criterion for model selection was Akaike's (1973) entropic information criterion (AIC). AIC is an information-theoretic criterion for model selection which was developed by Akaike (1973, 1974, 1976, 1977, 1978, 1979) in a series of papers. Given a discrete set of competing models indexed by k, $\{M_k : \ k = 1, 2, \ldots, K\}$, the AIC for a fixed model with k parameters is usually defined as -2 times the log likelihood evaluated at its corresponding Maximum Likelihood Estimators (MLEs) plus $2k$, that is,

$$AIC = -2 \log L_{M_k}(\hat{\theta}_k | x) + 2k.$$

The format and style follow that of *Journal of the American Statistical Association.*

The AIC procedure is to select the model which minimizes the AIC. Although, the AIC procedure is simple and versatile, several authors, including Bhansali and Downham (1977) and Schwarz (1978), have shown that the AIC procedure does not necessarily yield an asymptotically consistent estimator of the model order. This fact has prompted the use of other model selection criteria.

Schwarz (1978) approached the problem of model selection from a Bayesian perspective. He considered only independent and identically distributed (iid) observations from a probability distribution in the regular exponential family. Given a discrete set of competing models indexed by k, $\{M_k : \ k = 1, 2, \ldots, K\}$, the Schwarz Information Criterion or SIC for a fixed model with k parameters is defined as

$$SIC = \log L_{M_k}(\hat{\theta}_k|x) - \frac{1}{2} k \log(n),$$

where n is the total number of observations in the dataset of interest. Schwarz showed, by ad hoc methods, that for iid observations from a regular exponential family and for a broad range of possible priors, SIC is an approximation to the log of the posterior probability of model M_k. Thus the SIC procedure, choosing the model which maximizes SIC, approximately corresponds to the Bayesian procedure of selecting the model with greatest posterior probability.

In order to make Akaike's information-theoretic approach and Schwarz's Bayesian approach more similar, most recent work has focused on the Bayesian Information Criterion or BIC, which is proportional to Schwarz's SIC,

$$BIC = -2SIC = -2 \log L_{M_k}(\hat{\theta}_k|x) + k \log(n).$$

Bozdogan (1987) extended the AIC to the Consistent AIC (CAIC) using entropic information ideas similar to Akaike's (1973). He obtained

$$CAIC = -2 \log L_{M_k}(\hat{\theta}_k|x) + k[(\log n) + 1],$$

which is also similar to BIC.

The BIC has also been extended and improved by a number of authors. One direction has been to extend the class of models for which BIC applies. Haughton (1988) extended the derivation of BIC to the curved exponential family while Cavanaugh and Neath (1999) derived a generalization of BIC that only requires that the likelihood function satisfies certain regularity conditions without having the likelihood assuming a specific form.

Another direction in which BIC has been extended is to obtain even more accurate approximations to the posterior probability by carrying out the approximations more carefully. Kashyap (1982) noted that Schwarz's approximation to posterior probability can be viewed as a special case of Laplace's method and as a result gives a more accurate approximation including the constant terms in the expansion. Kass and Wasserman (1995) noted that BIC is more directly related to the log of the Bayes factor than to the log of the posterior probabilities.

The Bayes factor is defined as the ratio of posterior odds and prior odds. Since posterior odds are a simple transformation of Bayesian posterior probabilities, the Bayes factor can be used to select a model from a discrete set of competing models. Kass and Raftery (1995) provided a review of Bayes factors.

Evaluating Bayes factors involves the computation of posterior integrals and exact analytical evaluation does not always exist. Several analytical approximations to posterior integrals exist, one of which is Laplace's method. Laplace's (1820 or 1847) method is commonly used to provide an analytical approximation to integrals that take a particular form. A recent review of asymptotic expansion of integrals including Laplace's method is given by Olver (1997). Early uses of Laplace's method in Bayesian statistics were typically in theoretical contexts. Various authors including LeCam (1953, 1956, 1958, 1966) contributed to the literature by providing the regularity conditions required in using Laplace's method while other authors such as Johnson (1967, 1970) provided work on higher-order expansion approximations. The availability of affordable and fast computing has expanded the popularity of the use of Laplace's method in a more computational context in statistics. Tierney and Kadane (1986) provided an early use of Laplace's method in such a context. The two authors wrote the posterior expected value of a real-valued function of interest, $E[g(\theta)|X]$, as a ratio of two integrals and applied Laplace's method to both the numerator and denominator of that ratio. Tierney and Kadane showed that the resulting approximation has good accuracy since the leading terms of the approximation errors in the numerator and denominator cancel out. Tierney, Kass and Kadane (1987, 1989a, 1989b), Kass, Tierney and Kadane (1988, 1989, 1990, 1991), Wong and Li (1992) worked on the extensions to this methodology. More recently, Laplace approximation was used to approximate Bayes factors for nested models by Kass and Vaidyanathan (1992), for generalized linear models by Raftery (1996), and for variance component models by Pauler, Wakefield, and Kass (1999).

Kass and Wasserman (1995) studied BIC as an approximation to the log of the Bayes factor for iid observations. They showed that for a particular class of reference priors the log of the Bayes factor is approximated by BIC with error of order $O_p(n^{-1/2})$ instead of the more typical error of order $O_p(1)$. However in many statistical applications the observations are neither independent nor identically distributed. We will show that in the case where the observations are not iid, BIC might not provide a good approximation to the log of the Bayes factor and an alternative to BIC is needed as a model selection criterion. We will further show that while the first order approximations have been used in applying Laplace's method in approximating posterior integrals (see Kass and Vaidyanathan (1992) and Raftery (1996)), there are linear models where straightforward first order approximations will not work.

It is not a standard statistical practice to provide a justification prior to using Laplace's method to approximate a posterior integral. As discussed in Kass, Tierney and Kadane (1990), it is a common oversight to presume without any justification that the regularity conditions that rigorously justify applying Laplace's method hold in one's analysis. Moreover, the necessary regularity conditions such as those derived in Kass, Tierney and Kadane (1990) and Johnson (1967), are typically derived for iid observations. Given the non-iid nature of our model's covariance structure, it is not completely clear that these regularity conditions hold, and in fact there is one case where they fail to hold. Thus we find it necessary to justify the use of Laplace's method in approximating our posterior integrals.

Laplace's method was recently applied to variance component models by Pauler, Wakefield, and Kass (1999). While this paper is applicable to many variance component models, we are unable to apply the results from this paper to our model since we cannot assume that the cubic term in our asymptotic expansion is negligible at the boundary.

This research will consist of two parts. The first will be to look at the posterior probability integrals for a family of linear models which arises from work of Hart, Koen and Lombard (2003). Applying Laplace's method to these integrals is not entirely straightforward. One of the requirements is to analyze the asymptotic behavior of the information matrices as the sample size tends to infinity. This requires a number of analytic tricks, including viewing our covariance matrices as tending to differential operators. The use of differential operators and their Green's functions can provide a convenient and systematic method to asymptotically invert the covariance matrices. Once we have found the asymptotic behavior of the information matrices, we will see that in most cases BIC provides a reasonable approximation to the log of the posterior probability and Laplace's method gives more terms in the expansion and hence provides a slightly better approximation. In other cases, a number of pathologies will arise. We will see that in one case, BIC does not provide an asymptotically consistent estimate of the posterior probability; however, the more general Laplace's method will provide such an estimate. In another case, we will see that a naive application of Laplace's method will give a misleading answer and Laplace's method must be adapted to give the correct answer. The second part will be to use numerical methods to compute the "exact" posterior probabilities and compare them to the approximations arising from BIC and Laplace's method.

CHAPTER II

LAPLACE'S METHOD

A large fraction of this thesis will involve applying Laplace's method to approximate integrals. Laplace's method is a very general method for approximating integrals of the form

$$\int_{\Theta} b(\theta) e^{h(\theta,n)} d\theta \tag{2.1}$$

as n tends to infinity. Here Θ is a subset of Euclidean space $\mathbf{R}^p$, $b(\theta)$ is a function of θ alone, and $h(\theta, n)$ is a function of both θ and n. In our applications Θ will be the parameter space of a model, $b(\theta)$ will usually be the prior density on the parameters θ and $h(\theta, n)$ will usually be the log-likelihood of the parameters. The intuition one should have is that as n tends to infinity h becomes more rapidly varying while b, which does not depend on n, remains smooth. Because of the magnifying effect of the exponential, the dominant contribution to the integral comes from near the maximum of h. The contribution from near the maximum can then be approximated by replacing b by its value at this maximum and h by its quadratic Taylor polynomial. We will not attempt to give the most general possible formulation of Laplace's method. For more generality see Olver (1997). We will be content with generality sufficient for our applications.

The first condition we need to make this intuition precise is that $h(\theta, n)$ should have a unique global maximum $\hat{\theta}_n$ in the interior of Θ, at least for large n. In fact, we need a little more. This maximum must give the dominant contribution to the integral and must not be too close to the boundary of Θ. We will return to these two conditions later when we have more notation. We need $b(\theta)$ to be sufficiently smooth. For this, we assume $b(\theta)$ is continuously differentiable, bounded and positive on Θ. Further assume the first-order partial derivatives of $b(\theta)$ are bounded on Θ. We will also assume $b(\theta)$ is integrable on Θ. This condition together with the unique maximum of $h(\theta, n)$ ensures existence of the integral we are trying to approximate and holds in our examples, though much weaker conditions would suffice. We need $h(\theta, n)$ to be well approximated by its quadratic Taylor polynomial. For this, we first insist that $h(\theta, n)$ be thrice continuously differentiable as a function of θ on Θ. It follows that $\nabla h(\hat{\theta}_n, n) = 0$ and the quadratic Taylor approximation is

$$h(\theta, n) = h(\hat{\theta}_n, n) - \frac{1}{2}(\theta - \hat{\theta}_n)' H_n (\theta - \hat{\theta}_n) + R_n(\theta),$$

where

$$H_n = -\frac{\partial^2}{\partial \theta^2}\, h(\theta, n)|_{\theta = \hat{\theta}_n}$$

and $R_n(\theta)$ is the remainder term. Since $h(\theta, n)$ has a local maximum at $\hat{\theta}_n$, the Hessian matrix H_n must be positive semidefinite. In order for the quadratic approximation to be good, we need H_n to be positive definite, but this alone is not enough. We want $h(\theta, n)$ to fall off rapidly away from $\hat{\theta}_n$. Let λ_n be the smallest eigenvector of H_n. Then we need $\lambda_n \to \infty$ as $n \to \infty$. We will need additional conditions to guarantee the quadratic approximation is accurate.

We can now say more precisely the conditions we need from the global maximum. We want any contributions to the integral away from $\hat{\theta}_n$ to be negligible. One way to enforce the dominance which is sufficient in our applications, is to suppose that there is a sequence (a_n) such that $a_n / \log(n) \to \infty$ as $n \to \infty$, $\lambda_n / a_n \to \infty$ as $n \to \infty$ and

$$h(\hat{\theta}_n, n) - h(\theta, n) < a_n/3 \Rightarrow (\theta - \hat{\theta}_n)' H_n (\theta - \hat{\theta}_n) < a_n.$$

This condition says that choices of θ which give almost as large a value of h as the maximum are in fact close to the maximum. Let

$$\Theta_{0,n} = \{\theta :\ (\theta - \hat{\theta}_n)' H_n (\theta - \hat{\theta}_n) < a_n\}.$$

Then $\Theta_{0,n}$ are the only points we need to consider. The condition that the maximum not be too near the boundary just requires that these points all be in Θ. The quadratic approximation needs to be accurate only for these points. Thus

$$s_n = \sup_{\theta \in \Theta_{0,n}} \frac{|R_n(\theta)|}{(\theta - \hat{\theta}_n)' H_n (\theta - \hat{\theta}_n)} \longrightarrow 0 \text{ as } n \to \infty.$$

Summarizing the discussion above the required technical conditions are as follows:

1. $b(\theta)$ is continuously differentiable, bounded, positive and integrable on Θ. Further the first-order partial derivatives of $b(\theta)$ are bounded on Θ.

2. $h(\theta, n)$ has a unique global maximum $\hat{\theta}_n$ in the interior of Θ.

3. Let $H_n = -\frac{\partial^2}{\partial \theta^2} h(\theta, n)|_{\hat{\theta}_n}$ be the Hessian matrix. Then H_n should be positive definite and the smallest eigenvalue λ_n of H_n should tend to infinity as $n \to \infty$.

4. There is a sequence (a_n) such that $a_n / \log(n) \to \infty$ as $n \to \infty$, $\lambda_n / a_n \to \infty$ as $n \to \infty$ and

$$h(\hat{\theta}_n, n) - h(\theta, n) < a_n/3 \Rightarrow (\theta - \hat{\theta}_n)' H_n (\theta - \hat{\theta}_n) < a_n.$$

5.

$$\Theta_{0,n} = \{\theta : \ (\theta - \hat{\theta}_n)' H_n (\theta - \hat{\theta}_n) < a_n\} \subset \Theta.$$

6.

$$s_n = \sup_{\theta \in \Theta_{0,n}} \frac{|R_n(\theta)|}{(\theta - \hat{\theta}_n)' H_n (\theta - \hat{\theta}_n)} \longrightarrow 0 \text{ as } n \to \infty.$$

Suppose these technical conditions hold. Then we first split the integral into the portion over $\Theta_{0,n}$ and the portion over the complement.

$$\int_{\Theta} b(\theta) e^{h(\theta,n)} d\theta = \int_{\Theta_{0,n}} b(\theta) e^{h(\theta,n)} d\theta + \int_{\Theta - \Theta_{0,n}} b(\theta) e^{h(\theta,n)} d\theta.$$

Since $b(\theta)$ is integrable, say with $\int_\Theta b(\theta)d\theta = M$, and on $\Theta - \Theta_{0,n}$ we have $h(\hat{\theta}_n, n) - h(\theta, n) \geq a_n/3$, the second integral is at most

$$I_2 \leq M e^{h(\hat{\theta}_n,n)} e^{-a_n/3}. \tag{2.2}$$

Since $a_n/\log(n) \to \infty$, we will see that this second integral is negligible compared to the first. We thus turn to approximating the first integral, I_1. Since $\lambda_n/a_n \to \infty$ as $n \to \infty$, we see that

$$\Theta_{0,n} \subset \{\theta : \ \|\theta - \hat{\theta}_n\| \leq (a_n/\lambda_n)^{1/2}\},$$

i.e., $\Theta_{0,n}$ is contained in a small ball about $\hat{\theta}_n$. Since the partial derivatives of $b(\theta)$ are bounded and $b(\hat{\theta}_n) > 0$, we make only a relative error of $O((a_n/\lambda_n)^{1/2})$ in replacing $b(\theta)$ with $b(\hat{\theta}_n)$. The result is

$$I_1 = b(\hat{\theta}_n) \int_{\Theta_{0,n}} e^{h(\theta,n)} \, d\theta \left(1 + O((a_n/\lambda_n)^{1/2})\right).$$

From the Taylor approximation we have

$$h(\hat{\theta}_n, n) - \frac{1}{2}(1+2s_n)(\theta-\hat{\theta}_n)' H_n(\theta-\hat{\theta}_n) \leq h(\theta, n) \leq h(\hat{\theta}_n, n) - \frac{1}{2}(1-2s_n)(\theta-\hat{\theta}_n)' H_n(\theta-\hat{\theta}_n).$$

Thus

$$b(\hat{\theta}_n) e^{h(\hat{\theta}_n,n)} \int_{\Theta_{0,n}} \exp\left(-\frac{1}{2}(1+2s_n)(\theta-\hat{\theta}_n)' H_n(\theta-\hat{\theta}_n)\right) \, d\theta \left(1 + O((a_n/\lambda_n)^{1/2})\right) \leq I_1$$
$$\leq b(\hat{\theta}_n) e^{h(\hat{\theta}_n,n)} \int_{\Theta_{0,n}} \exp\left(-\frac{1}{2}(1-2s_n)(\theta-\hat{\theta}_n)' H_n(\theta-\hat{\theta}_n)\right) \, d\theta \left(1 + O((a_n/\lambda_n)^{1/2})\right).$$

Making the substitutions $\mathbf{t} = (1 \pm s_n)^{1/2} H_n^{1/2}(\theta - \hat{\theta}_n)$ in the two bounding integrals gives

$$(1+2s_n)^{-p/2} \det(H_n)^{-1/2} b(\hat{\theta}_n) e^{h(\hat{\theta}_n,n)} \int_{\{\mathbf{t}\in\mathbf{R}^p : \|\mathbf{t}\|^2 \leq a_n(1+2s_n)\}} \exp\left(-\frac{1}{2}\|\mathbf{t}\|^2\right) d\mathbf{t} \times$$
$$\left(1 + O((a_n/\lambda_n)^{1/2})\right) \leq I_1 \leq (1-2s_n)^{-p/2} \det(H_n)^{-1/2} b(\hat{\theta}_n) e^{h(\hat{\theta}_n,n)} \times$$
$$\int_{\{\mathbf{t}\in\mathbf{R}^p : \|\mathbf{t}\|^2 \leq a_n(1-2s_n)\}} \exp\left(-\frac{1}{2}\|\mathbf{t}\|^2\right) d\mathbf{t} \left(1 + O((a_n/\lambda_n)^{1/2})\right).$$

The upper bound will only get larger if we extend the region of integration to be all of $\mathbf{R}^p$. For the lower bound, the relative error in extending the integration to be over all of $\mathbf{R}^p$ is $O(a_n^{(p-2)/2}\exp(-a_n/2))$. Thus we get

$$(2\pi)^{p/2}(1+2s_n)^{-p/2}\det(H_n)^{-1/2}b(\hat\theta_n)e^{h(\hat\theta_n,n)}\left(1+O((a_n/\lambda_n)^{1/2}, a_n^{(p-2)/2}\exp(-a_n/2))\right)$$
$$\le I_1 \le (2\pi)^{p/2}(1-2s_n)^{-p/2}\det(H_n)^{-1/2}b(\hat\theta_n)e^{h(\hat\theta_n,n)}\left(1+O((a_n/\lambda_n)^{1/2})\right).$$

Combining these two bounds and recalling that $s_n \to 0$ as $n \to \infty$ gives

$$I_1 = (2\pi)^{p/2}\det(H_n)^{-1/2}b(\hat\theta_n)e^{h(\hat\theta_n,n)}\left(1+O(s_n, (a_n/\lambda_n)^{1/2}, a_n^{(p-2)/2}\exp(-a_n/2))\right).$$

Combining this with the bound on I_2 from (2.2) gives

$$\int_\Theta b(\theta)e^{h(\theta,n)}d\theta = (2\pi)^{p/2}\det(H_n)^{-1/2}b(\hat\theta_n)e^{h(\hat\theta_n,n)}\times$$
$$\left(1+O(s_n, (a_n/\lambda_n)^{1/2}, a_n^{(p-2)/2}\exp(-a_n/2), \det(H_n)^{1/2}\exp(-a_n/3))\right).$$

In our examples, the Hessian matrix H_n and the residual $R_n(\theta)$ will scale like powers of the sample size as n tends to infinity. For example, if $h(\theta)$ is the log-likelihood for an i.i.d. sample from an exponential family, then H_n and R_n both scale roughly like n for large n. Thus $\lambda_n = O(n^{-1})$ and $s_n = O(n^{-1})$. Since $a_n/\log(n) \to \infty$, the last two terms in the error are negligible and the error is $O((\log(n)/n)^{1/2})$. If we assume more differentiablity for $b(\theta)$ and $h(\theta)$, then the we can carry more terms in the Taylor expansion of b and h in the calculation above. The first order corrections give integrals of odd powers of $\theta - \hat\theta$ times the Gaussian and therefore vanish. This gives stronger bounds on the error terms.

CHAPTER III

MODEL FORMULATION AND DEFINITIONS

Hart, Koen and Lombard (2003) recently proposed a linear model in their analysis of variable stars. The lengths of time between successive maximum brightnesses of a variable star will be referred to as pseudo-periods. These pseudo-periods tend to fluctuate substantially about the star's actual period, which is the long-run average length of time between successive maximum brightnesses. Periods of variable stars are meaningful to astronomers as many characteristics of these stars may be studied from their periods. For a single star the data are (j, Y_j), $j = 1, \ldots, n$, where n is the sample size of that particular star and $Y_1, Y_2, \ldots, Y_n$ are the measured pseudo-periods of the star. They are treated as a time series and are modeled as follows:

$$Y_j = \mu_m(j) + I_j + \epsilon_j - \epsilon_{j-1}, \quad j = 1, \ldots, n,$$

where μ_m is an mth degree polynomial that is used to model the systematic trend of each star, $m = 0, \ldots, 15$. The true degree of the polynomial is unknown and is one of the quantities we wish to determine through our model selection. The quantity I_j corresponds to the random variation intrinsic to the star, and ϵ_j is the experimental error made in measuring the jth time of maximum brightness.

The ϵ_j's are assumed to be independent normal random variables with mean 0 and finite variance. The ϵ_j's are allowed to be heteroscedastic in the following manner:

$$\mathrm{Var}(\epsilon_j) = \exp(2(\beta_0 + \beta_1 j)), \quad j = 1, \ldots, n.$$

The justification for this model is that an earlier analysis indicated a tendency for residual variance to decrease over time, which is consistent with the fact that measurement methods have improved over time. The I_j's are modeled as a first order autoregressive, AR(1), process, i.e.,

$$I_j = \rho I_{j-1} + Z_j, \quad j = 2, \ldots, n,$$

where $|\rho| < 1$ and the $Z_2, \ldots, Z_n$ are iid normal random variables with mean 0 and finite variance σ_Z^2. Let $\sigma_I^2 = \mathrm{Var}(I_j) = \sigma_Z^2/(1-\rho^2)$. For many of the stars we will see that the data suggests that $\sigma_I^2 = 0$, i.e., that the covariance structure is given entirely by the ϵ_j's. The presence or absence of the I_j part of the covariance structure is the second part of the model selection we wish to perform. Thus a model for us will be described by a pair $M = (m, h)$ where m is the degree of the polynomial fitted to the means and h is an indicator with $h = 0$ indicating $\sigma_I^2 = 0$ and $h = 1$ indicating $\sigma_I^2 > 0$.

The Y_j's are thus distributed multivariate normal with means

$$E(Y_j) = \theta_0 + \theta_1 \frac{j}{n} + \cdots + \theta_m \left(\frac{j}{n}\right)^m, \quad j = 1, \ldots, n. \tag{3.1}$$

We will consider the models with $0 \le m \le 15$. Let Θ_m denote the parameter space of θ_m for the degree m model. The covariance matrix Σ of $Y_1, \ldots, Y_n$ is given by

$$\mathrm{Cov}(Y_i, Y_j) = \begin{cases} \frac{\sigma_Z^2}{1-\rho^2} + \exp(2(\beta_0 + \beta_1 j)) + \exp(2(\beta_0 + \beta_1(j-1))), & i = j \\ \frac{\rho\sigma_Z^2}{1-\rho^2} - \exp(2(\beta_0 + \beta_1 \min(i,j))), & |i-j| = 1 \\ \frac{\rho^{|i-j|}\sigma_Z^2}{1-\rho^2}, & |i-j| > 1. \end{cases} \tag{3.2}$$

Let η denote the covariance parameters for the model, either (β_0, β_1) if $h = 0$ or $(\sigma_I^2, \rho, \beta_0, \beta_1)$ if $h = 1$. Let Ω_h denote the parameter space for η for the model with indicator h. For brevity we will omit the subscript h whenever it is immaterial.

Considered as a function of the parameters the likelihood can be written as

$$f(\mathbf{Y}|\mu_m, \Sigma) = \frac{1}{(2\pi)^{n/2}} (\det(\Sigma))^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{Y} - \mu_m)'\Sigma^{-1}(\mathbf{Y} - \mu_m)\right).$$

where $\mu_m = \mu_m(\theta_m)$ and $\Sigma = \Sigma_h(\eta)$ are given by formulas (3.1) and (3.2), respectively.

Let α_M denote the prior probability of model M. Assume that the mean parameters θ_m and the covariance parameters are *a priori* independent. Then the prior has the form

$$\pi(\theta_m, \eta | M) = \pi_m(\theta_m)\pi_h(\eta).$$

Let $Z(\mathbf{Y})$ be the marginal density of $\mathbf{Y}$. The posterior probability of model M given the data is then

$$\pi(M|\mathbf{Y}) = \frac{\alpha_M}{Z(\mathbf{Y})} \int_{\Omega_h} \int_{\Theta_m} f(\mathbf{Y}|\mu_m, \Sigma)\pi_m(\theta_m)\pi_h(\eta)\ d\theta_m d\eta$$

In evaluating whether BIC is a good approximation for this star model we use Laplace's method to approximate this posterior integral in the next chapter. In Chapter V, we provide an approximation to the exact posterior probabilities.

CHAPTER IV

ANALYTICAL APPROXIMATION TO THE POSTERIOR PROBABILITIES USING LAPLACE'S METHOD

From our description of the models of interest in Chapter III, we see that the data $\mathbf{Y}$ are normally distributed with mean $\mu_m(j)$, where $\mu_m(j) = X\theta_m$, and variance Σ. Note that we have supressed the dependence of the design matrix X on m and n from the notation. The posterior probability of model M given the data is

$$\pi(M|\mathbf{Y}) = \frac{\alpha_M}{Z(\mathbf{Y})(2\pi)^{n/2}} \int_{\Omega_h} (\det(\Sigma))^{-1/2} \pi_h(\eta) \times \\ \int_{\Theta_m} \exp\left(-\frac{1}{2}(\mathbf{Y} - X\theta_m)'\Sigma^{-1}(\mathbf{Y} - X\theta_m)\right) \pi_m(\theta_m) d\theta_m d\eta.$$

Since our data $\mathbf{Y}$ are normally distributed, the parameters θ_m which only influence the mean of $\mathbf{Y}$ and the parameters η which only influence the covariance matrix of $\mathbf{Y}$ are null orthogonal (and in fact orthogonal). Let $\hat{\Sigma}$, $\hat{\eta}$ and $\hat{\theta}_m$ denote the MLEs for these quantities for model M. Note that

$$\hat{\theta}_m = (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}\mathbf{Y}$$

is also a generalized least squares estimator of θ_m. The information matrix for η is given by

$$I_{\eta,\eta} = \left(\frac{1}{2}\operatorname{tr}\left(\Sigma^{-1}\frac{\partial\Sigma}{\partial\eta_i}\Sigma^{-1}\frac{\partial\Sigma}{\partial\eta_j}\right)\right)$$

and the information matrix for θ_m is given by

$$I_{\theta,\theta} = X'\Sigma^{-1}X.$$

Assume for now that the hypotheses necessary for the Laplace approximation hold, then the resulting approximation to the posterior probability will be

$$\pi(M|\mathbf{Y}) = \frac{\alpha_M}{Z(\mathbf{Y})(2\pi)^{n/2}}(\det\hat{\Sigma})^{-1/2}\exp(-\frac{1}{2}(\mathbf{Y}-X\hat{\theta}_m)'\hat{\Sigma}^{-1}(\mathbf{Y}-X\hat{\theta}_m))\times$$
$$\pi_m(\hat{\theta}_m)\pi_h(\hat{\eta})(2\pi)^{(m+3+2h)/2}(\det I_{\hat{\eta},\hat{\eta}})^{-1/2}(\det I_{\hat{\theta},\hat{\theta}})^{-1/2}\left(1+O_p(1/n)\right). \tag{4.1}$$

Chapter II gives the technical conditions which must be met for the Laplace approximation given in formula (4.1) to be valid. These conditions require that we understand the asymptotic behavior of the information matrices $I_{\eta,\eta}$ and $I_{\theta,\theta}$ as n tends to infinity. In particular, for the Laplace approximation to be valid we need all eigenvalues of these matrices to tend to infinity as n tends to infinity. Further, the asymptotic formulas for these matrices will be part of the asymptotic formulas for the posterior probability.

4.1 The Information Matrix of θ

In order to study the asymptotic behavior of $I_{\theta,\theta}$, we will take the following approach. Any vector $\mathbf{v} = (v_1, v_2, \ldots, v_n)$, such as a column of the design matrix X, can be viewed as a step function on $(0,1]$ by identifying $\mathbf{v}$ with the function $f_{\mathbf{v}}(t) = v_i$ for $(i-1)/n < t \leq i/n$, $i = 1,\ldots,n$. With this interpretation the dot product is interpreted as integration via the formula

$$\mathbf{v}\cdot\mathbf{w} = n\int_0^1 f_{\mathbf{v}}(t)f_{\mathbf{w}}(t)dt.$$

This interpretation makes it possible to identify the limit of a sequence of vectors of length n as n tends to infinity with a function on $[0,1]$. For example, for our design matrix X the ith column is $X_i = ((1/n)^i, (2/n)^i, \ldots, (n-1/n)^i, (n/n)^i)$ and $\lim_{n\to\infty} f_{X_i}(t) = t^i$. For most regression models one would care to write down, including polynomial models such as the present case or Fourier series, the columns of the design matrix will have a nice limiting behavior.

Similarly if $A = (a_{i,j})$ is an $n \times n$ matrix, then we can interpret A as a piecewise constant function $a(s,t)$ on $(0,1] \times (0,1]$ by setting $a(s,t) = a_{i,j}$ for $(i-1)/n < s \leq i/n$ and $(j-1)/n < t \leq j/n$. Hence matrix multiplication becomes integration as well. Specifically, if the matrix A corresponds to the function $a(s,t)$, the matrix B corresponds to $b(s,t)$, and the vector $\mathbf{v}$ corresponds to the function $f_{\mathbf{v}}(t)$, then the vector $A\mathbf{v}$ corresponds to

$$f_{A\mathbf{v}}(s) = n \int_0^1 a(s,t) f_{\mathbf{v}}(t) dt$$

and the matrix AB corresponds to the function

$$n \int_0^1 a(s,\tau) b(\tau,t) d\tau.$$

As before this interpretation makes it possible to identify the limit of a sequence of $n \times n$ matrices as n tends to infinity with the limit of the corresponding functions. Unfortunately, for most of the covariance matrices we are interested in, taking this limit will require rescaling by a power of n and the limit will need to be interpreted as a distribution on $[0,1] \times [0,1]$ rather than as a function. For example, the $n \times n$ identity matrix I_n corresponds to the function $i_n(s,t)$ which is 1 if $s,t \in ((i-1)/n, i/n]$ for some i and zero otherwise. Hence $\lim_{n\to\infty} n i_n(s,t) = \delta(s-t)$ where δ denotes the Kronecker delta function. We will abbreviate this by saying that I_n corresponds to $\delta(s-t)/n + O(n^{-2})$.

For the covariance structure of our model the covariance matrix $\Sigma = A + B$ can be broken into the two parts A and B. Here A represents the part of the covariance matrix of the star model coming from the ϵ_j's and B represents the part of the covariance matrix coming from the I_j's. Specifically, we have $A = (a_{i,j})$ where

$$a_{i,j} = \begin{cases} \exp(2(\beta_0 + \beta_1 j)) + \exp(2(\beta_0 + \beta_1 (j-1))), & i = j \\ -\exp(2(\beta_0 + \beta_1 \min(i,j))), & |i-j| = 1 \\ 0, & |i-j| > 1. \end{cases} \tag{4.2}$$

The second part $B = (b_{i,j})$ of the covariance matrix Σ is given by

$$b_{i,j} = \sigma_I^2 \rho^{|i-j|}.$$

The matrix A can be inverted explicitly to give

$$(A^{-1})_{i,j} = \frac{e^{\beta_1 - 2\beta_0}}{e^{\beta_1} - e^{-\beta_1}} \cdot \frac{e^{2(n+1-\max(i,j))\beta_1} - e^{2(n+1-i-j)\beta_1} - 1 + e^{-2\min(i,j)\beta_1}}{e^{2(n+1)\beta_1} - 1}.$$

This formula has a removable singularity at $\beta_1 = 0$, the correct formula in that case can be obtained by taking the limit as β_1 tends to zero. The parameter β_0 which relates to the variance of the first measurement error ϵ_0 should not depend on the number of observations n. However β_1 should scale like $\beta_1 = b/n$ for some constant b. Otherwise the variances will change dramatically between the first and last observation and as a result only a small fraction of the observations will actually contribute to our parameter estimates. This scaling seems to be borne out by the data. For the scaling $\beta_1 = b/n$, one can show that $(1/n)A^{-1}$ converges as n tends to infinity to the function

$$g(s,t) = \frac{e^{-2\beta_0}}{2b(e^{2b} - 1)} \left(e^{2b(1-\max(s,t))} - e^{2b(1-s-t)} - 1 + e^{-2b\min(s,t)}\right). \tag{4.3}$$

Alternately, (4.3) can be derived without explicitly inverting A using techniques which would be helpful for a large number of covariance structures. Consider multiplying the matrices A by a sequence of vectors $\mathbf{v}$ which converge to the smooth function $f_{\mathbf{v}}(t)$. Then, ignoring boundary effects or assuming $f_{\mathbf{v}}(0) = f_{\mathbf{v}}(1) = 0$, we compute

$$\lim_{n\to\infty} n^2 f_{A\mathbf{v}}(t) = -e^{2\beta_0} \frac{d}{dt}\left(e^{2bt} \frac{df_{\mathbf{v}}(t)}{dt}\right). \tag{4.4}$$

Thus $\lim_{n\to\infty} n^2 A$ can be interpreted as a differential operator. The inverse to a differential operator will be the corresponding Green's function. Specifically, suppose we have a sequence of vectors $\mathbf{w}$ converging to $f_{\mathbf{w}}(t)$. Since $(1/n)A^{-1}$ converges to $g(s,t)$, $(1/n^2)A^{-1}w$ will converge to

$$h(t) = \int_0^1 g(t,\tau) f_{\mathbf{w}}(\tau) d\tau. \tag{4.5}$$

Hence by the identification of A with a differential operator in (4.4), we see that $n^2 A(1/n^2)A^{-1}w = w$ will converge to

$$-e^{2\beta_0} \frac{d}{dt}\left(e^{2bt} \frac{dh(t)}{dt}\right) = f_{\mathbf{w}}(t). \tag{4.6}$$

Combining (4.5) and (4.6) gives

$$-e^{2\beta_0} \frac{\partial}{\partial s}\left(e^{2bs} \frac{\partial g(s,t)}{\partial s}\right) = \delta(s-t).$$

Thus, $g(s,t)$ is the Green's function for the differential operator corresponding to $n^2 A$. Conversely we could have used this method to find the asymptotic behavior of A^{-1}. We first identify $n^2 A$ with the differential operator using (4.4), then directly compute the Green's function $g(s,t)$ for this differential operator on $[0,1]$ with the boundary conditions $g(0,t) = g(1,t) = 0$. Thus it follows that $(1/n)A^{-1}$ converges to this Green's function.

The second part $B = (b_{i,j})$ of the covariance matrix Σ is given by

$$b_{i,j} = \sigma_I^2 \rho^{|i-j|}.$$

Based on the data it appears that ρ does not scale with n and therefore the entries of this matrix fall off rapidly as they move away from the diagonal. For vectors which tend to smooth functions (the only type we will need to consider), entries near the diagonal have almost the same effect as diagonal entries (with errors of order n^{-1}).

Ignoring boundary effects which are also $O(1/n)$, the row sums of B are

$$\sum_{j=-\infty}^{\infty} \rho^{|j|} = \frac{1+\rho}{1-\rho}.$$

Suppose the vector $\mathbf{v}$ represents a smooth function f in the sense that $\mathbf{v}' = (f(1/n), f(2/n), \ldots, f(1))$. Then we have

$$\left\| (B - \sigma_I^2 \, \frac{1+\rho}{1-\rho} \, I_n)\mathbf{v} \right\| \leq O(\|\mathbf{v}\|/n).$$

We will abbreviate this as

$$B = \sigma_I^2 \, \frac{1+\rho}{1-\rho} \, I_n(1 + O(1/n)). \tag{4.7}$$

Hence if $\sigma_I^2 > 0$ we have

$$B^{-1} = \sigma_I^{-2} \, \frac{1-\rho}{1+\rho} \, I_n(1 + O(1/n)). \tag{4.8}$$

When considering $I_{\theta,\theta} = X'\Sigma^{-1}X$, the natural measure of the size of a matrix A is the matrix norm $\max_{\{\mathbf{v}:\|\mathbf{v}\|\neq 0\}} \|A\mathbf{v}\|/\|\mathbf{v}\|$ where the maximum is taken over vectors $\mathbf{v}$ which represent smooth functions. Thus the matrix A of (4.2) has size $O(n^{-2})$, A^{-1} has size $O(n^2)$, and provided $\sigma_I^2 > 0$, B and B^{-1} have size $O(1)$. Thus if $\sigma_I^2 > 0$, A is much smaller than B and $\Sigma \approx B = O(1)$, but if $\sigma_I^2 = 0$, then $\Sigma = A = O(n^{-2})$. This difference in scales will result in differences in the Laplace approximations.

First suppose $\sigma_I^2 > 0$. The columns of the X matrix converge to functions on $[0, 1]$; therefore X will converge to a row vector of functions:

$$\lim_{n\to\infty} X = (f_0(t) \; f_1(t) \; \ldots \; f_m(t)).$$

Since

$$\Sigma = B + O(n^{-2}) = \sigma_I^2 \, \frac{1+\rho}{1-\rho} \, I_n(1 + O(1/n)),$$

we have

$$\lim_{n\to\infty} n\Sigma^{-1} = \sigma_I^{-2} \, \frac{1-\rho}{1+\rho} \, \delta(s-t)$$

and hence for $0 \le i, j \le m$ we have

$$(X'\Sigma^{-1}X)_{i,j} \sim n^2 \int_0^1 \int_0^1 f_i(s) n^{-1} \sigma_I^{-2} \, \frac{1-\rho}{1+\rho} \, \delta(s-t) f_j(t) ds dt$$
$$= n\sigma_I^{-2} \frac{1-\rho}{1+\rho} \int_0^1 f_i(t) f_j(t) dt.$$

If the functions f_i are linearly independent on $[0, 1]$, as they are in our case and will be in any nonredundant regression model, then the matrix whose (i, j) entry is $\int_0^1 f_i(t) f_j(t) dt$ is positive definite. Since $X'\Sigma^{-1}X$ scales roughly like n times this matrix all the eigenvalues of $I_{\theta,\theta} = X'\Sigma^{-1}X$ will be large for large n, as required. Furthermore

$$\log\det(I_{\theta,\theta}) = (m+1)\log(n) + O(1)$$

which is consistent with the standard BIC formula. For our specific case of $f_i(t) = t^i$, we have

$$(X'\Sigma^{-1}X)_{i,j} \sim n\sigma_I^{-2} \, \frac{1-\rho}{1+\rho} \cdot \frac{1}{i+j+1}$$

and hence

$$\det(I_{\theta,\theta})^{-1/2} \approx n^{-(m+1)/2} \left(\frac{\sigma_I^2(1+\rho)}{1-\rho} \right)^{(m+1)/2} \prod_{i=1}^{m} (2i+1)^{1/2} \binom{2i}{i}.$$

Next suppose $\sigma_I^2 = 0$, in which case $\Sigma = A$. The X matrix is exactly as in the previous case but

$$\Sigma^{-1} = A^{-1} \sim n g(s, t)$$

where the Green's function $g(s, t)$ is given in (4.3). Hence for $0 \le i, j \le m$ we have

$$(X'\Sigma^{-1}X)_{i,j} \sim n^3 \int_0^1 \int_0^1 f_i(s) g(s, t) f_j(t) ds dt.$$

The matrices $L_m = (\ell_{i,j})_{0 \le i,j \le m}$ with

$$\ell_{i,j} = \int_0^1 \int_0^1 s^i g(s, t) t^j ds dt$$

are positive definite. To see this, let $p(t)$ be a nonzero polynomial and let $q(t) = \int_0^1 g(s,t)p(s)ds$ be the unique solution to $-e^{2\beta_0}\frac{d}{dt}(e^{2bt}q'(t)) = p(t)$ with $q(0) = q(1) = 0$. Then $q(t)$ is not constant and hence

$$\begin{aligned}\int_0^1 \int_0^1 p(s)g(s,t)p(t)dsdt &= \int_0^1 q(t)p(t)dt \\ &= -e^{2\beta_0}\int_0^1 q(t)\frac{d}{dt}e^{2bt}q'(t)dt \\ &= e^{2\beta_0}\int_0^1 e^{2bt}(q'(t))^2 dt > 0.\end{aligned}$$

It follows that all the eigenvalues of $I_{\theta,\theta}$ tend to infinity (but like n^3) as n tends to infinity. Hence

$$\log\det(I_{\theta,\theta}) = 3(m+1)\log n + \log\det(L_m) + O(1/n).$$

Note that this differs from the standard BIC formula. Because of the negative correlation between adjacent observations, we get more information about the regression coefficients θ_m than one might naively expect.

4.2 The Information Matrix of η

Next we consider the asymptotic behavior of $I_{\eta,\eta}$. Recall that for multivariate normal random variables, if the covariance matrix Σ depends on parameters $\eta = (\eta_1, \ldots, \eta_k)$, then the (i,j) entry of the information matrix is given by

$$I_{\eta_i,\eta_j} = \frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}\frac{\partial\Sigma}{\partial\eta_i}\Sigma^{-1}\frac{\partial\Sigma}{\partial\eta_j}\right) = -\frac{1}{2}\mathrm{tr}\left(\frac{\partial\Sigma^{-1}}{\partial\eta_i}\frac{\partial\Sigma}{\partial\eta_j}\right). \tag{4.9}$$

Consider the case where $\sigma_I^2 = 0$. This will be the case if the correct model has only the ϵ_j contribution to the error term, i.e., $h = 0$. Recall that even though the correct model may be the $h = 0$ model, the model whose posterior probability we are calculating may be any larger model. We will thus need the full 4×4 information matrix even if the truth is $h = 0$. The discussion in Section 4.1 of the asymptotic behavior of the covariance matrices A and B must be applied with care to this situation. The earlier discussion was for A and B or their inverses applied to vectors v which tend to a smooth function $f_{\mathbf{v}}$. The columns of $\Sigma^{-1} = A^{-1}$ after rescaling tend to continuously differentiable functions, but not twice differentiable functions. Thus we cannot expect to treat A as a second order differential operator. But the asymptotic behavior of B only required the function $f_{\mathbf{v}}$ to be Lipschitz continuous, therefore the discussion of Section 4.1 still applies. We saw in (4.7) that $B = \sigma_I^2 \, \frac{1+\rho}{1-\rho} \, I + O(n^{-1})$. Thus to leading order, the contribution of B will depend only on $\tau^2 = \sigma_I^2 \frac{1+\rho}{1-\rho}$. To capture this fact it is convenient to use $(\tau^2, \rho, \beta_0, b)$ as our parameters. Then

$$\frac{\partial \Sigma}{\partial \tau^2} = \frac{1-\rho}{1+\rho} \, (\rho^{|i-j|})_{1 \le i,j \le n} = I + O(1/n), \tag{4.10}$$

and

$$\frac{\partial \Sigma}{\partial \rho} = -\tau^2 \, \frac{2}{(1+\rho)^2} (\rho^{|i-j|})_{1 \le i,j \le n} + \tau^2 \, \frac{1-\rho}{1+\rho} (|i-j| \rho^{|i-j|-1})_{1 \le i,j \le n}. \tag{4.11}$$

Away from the boundaries (which are $O(1/n)$ corrections), the row sums of the matrix $\partial \Sigma / \partial \rho$ tend to zero. Hence, when applied to a sequence of vectors v which tend to a differentiable function $f_{\mathbf{v}}(t)$ we have

$$\frac{\partial \Sigma}{\partial \rho} v = O(\tau^2 / n). \tag{4.12}$$

For the parameter β_0 we have

$$\frac{\partial \Sigma}{\partial \beta_0} = 2A = 2\Sigma. \tag{4.13}$$

For the fourth parameter $b = n\beta_1$, note that since A is tridiagonal, so is $\frac{\partial \Sigma}{\partial b} = \frac{\partial A}{\partial b}$ and

$$\left(\frac{\partial \Sigma}{\partial b} \right)_{i,i} = \frac{2(i-1)\exp(2\beta_0 + 2(i-1)\beta_1) + 2i\exp(2\beta_0 + 2i\beta_1)}{n} \tag{4.14}$$

and

$$\left(\frac{\partial \Sigma}{\partial b}\right)_{i,i+1} = \left(\frac{\partial \Sigma}{\partial b}\right)_{i+1,i} = -\frac{2i\exp(2\beta_0 + 2i\beta_1)}{n}. \tag{4.15}$$

Also note the useful identity

$$A^{-1}_{i,i} - 2A^{-1}_{i,i+1} + A^{-1}_{i+1,i+1} = e^{-2\beta_0 - 2i\beta_1} - \frac{(1-e^{2\beta_1})e^{-2\beta_0 + 2(n-2i)\beta_1}}{e^{2(n+1)\beta_1} - 1}$$

which follows by direct computation. If $\sigma_I^2 = 0$, then $\tau^2 = 0$, $\partial\Sigma/\partial\rho = 0$ and all entries of the information matrix corresponding to ρ are of course zero. However we will want to apply this discussion to the case where σ_I^2, and hence τ^2, is small but positive. Thus we will need to compute the magnitude of the (ρ, ρ) entry of I in this case, though we will not need the off-diagonal ρ entries. Plugging in formulas (4.10 - 4.15) for the derivatives of Σ into the formula (4.9) for the information matrix gives:

$$\begin{aligned} I_{\tau^2,\tau^2} &\sim \frac{1}{2}\mathrm{tr}(A^{-2}) \sim \frac{n^4}{2}\int_0^1\int_0^1 (g(s,t))^2 ds dt \\ &= \frac{n^4}{2}\left(\frac{e^{4b} + e^{-4b} - 16e^{2b} - 16e^{-2b} + 30 + 48b^2}{192b^4(e^{2b}-1)^2}\right), \end{aligned} \tag{4.16}$$

$$\begin{aligned} I_{\tau^2,\beta_0} &\sim \mathrm{tr}(A^{-1}) \sim n^2\int_0^1 g(s,s)ds \\ &= n^2\left(\frac{e^{2b} - e^{-2b} - 4b}{8b^2(e^{2b}-1)}\right), \end{aligned} \tag{4.17}$$

$$\begin{aligned} I_{\tau^2,b} &\sim -\frac{1}{2}\mathrm{tr}\left(\frac{\partial A^{-1}}{\partial b}\right) \sim -\frac{n^2}{2}\int_0^1 \frac{\partial g(s,s)}{\partial b} ds \\ &= \frac{n^2}{2}\left(\frac{e^{6b} - (4b^2 + b + 1)e^{4b} - e^{2b} + 1 + b}{4b^3 e^{2b}(e^{2b}-1)^2}\right), \end{aligned} \tag{4.18}$$

$$\begin{aligned} I_{\rho,\rho} &= \frac{1}{2}\mathrm{tr}\left(A^{-1}\frac{\partial\Sigma}{\partial\rho}A^{-1}\frac{\partial\Sigma}{\partial\rho}\right) \\ &= O\left(\frac{\tau^4}{n^2}\mathrm{tr}(A^{-2})\right) \\ &= O(n^2\tau^4), \end{aligned} \tag{4.19}$$

$$I_{\beta_0,\beta_0} = \frac{1}{2}\mathrm{tr}(A^{-1}2AA^{-1}2A) = 2\mathrm{tr}(I) = 2n, \tag{4.20}$$

$$
\begin{aligned}
I_{\beta_0,b} &= \frac{1}{n}\mathrm{tr}\left(A^{-1}\frac{\partial A}{\partial \beta_1}\right) \\
&= \frac{e^{2\beta_0}}{n}\left\{\sum_{i=1}^{n} A_{i,i}^{-1}\left(2(i-1)e^{2(i-1)\beta_1}+2ie^{2i\beta_1}\right) - 2\sum_{i=1}^{n-1} A_{i,i+1}^{-1} 2ie^{2i\beta_1}\right\} \\
&= \frac{2e^{2\beta_0}}{n}\left\{ne^{2n\beta_1}A_{n,n}^{-1} + \sum_{i=1}^{n-1}(A_{i,i}^{-1} - 2A_{i,i+1}^{-1} + A_{i+1,i+1}^{-1})ie^{2i\beta_1}\right\} \\
&= \frac{2}{n}\left\{ne^{2n\beta_1}\frac{e^{2\beta_1}-e^{2(1-n)\beta_1}}{e^{2(n+1)\beta_1}-1} + \sum_{i=1}^{n-1}\left(i - \frac{i(1-e^{2\beta_1})e^{2(n-i)\beta_1}}{e^{2(n+1)\beta_1}-1}\right)\right\} \\
&= n + \frac{e^{2(n+1)\beta_1}+1}{e^{2(n+1)\beta_1}-1} - \frac{2(e^{2n\beta_1}-1)}{n(e^{2(n+1)\beta_1}-1)(1-e^{-2\beta_1})} \\
&= n + O(1),
\end{aligned}
\tag{4.21}
$$

and

$$
\begin{aligned}
I_{b,b} &= -\frac{1}{2n^2}\mathrm{tr}\left(\frac{\partial A^{-1}}{\partial b}\frac{\partial A}{\partial \beta_1}\right) \\
&= -\frac{e^{2\beta_0}}{n^2}\left\{ne^{2n\beta_1}\frac{\partial A_{n,n}^{-1}}{\partial b} + \sum_{i=1}^{n-1}\frac{\partial}{\partial b}\left(A_{i,i}^{-1} - 2A_{i,i+1}^{-1} + A_{i+1,i+1}^{-1}\right)ie^{2i\beta_1}\right\} \\
&= \frac{1}{n^2}\left\{O(n^2) + \sum_{i=1}^{n-1}(2i^2 + O(n))\right\} \\
&= \frac{2n}{3} + O(1).
\end{aligned}
\tag{4.22}
$$

Formulas (4.16)-(4.18) fail to make sense if $b = 0$ since both the numerator and denominator are zero. This apparent singularity is removable and the correct formula for this special case can be obtained by taking the limit as b tends to zero.

Suppose now that the true model is an $h = 0$ model and that the model whose posterior probability we are calculating is also an $h = 0$ model, possibly with higher degree. Then formulas (4.20), (4.21) and (4.22) show that the information matrix for η is

$$
I_{\eta,\eta} = n\begin{pmatrix} 2 & 1 \\ 1 & 2/3 \end{pmatrix} + O(1).
$$

All the eigenvalues of this matrix are large as n tends to infinity and therefore the likelihood will be sharply peaked about the MLEs with the dominant contribution to the posterior probability coming from η with $\|\eta - \hat{\eta}\| = O(n^{-1/2})$. Since

$$E\left(\frac{\partial^3 \log L(\eta|Y)}{\partial \eta_i^3}\right) = -\frac{3}{2}\mathrm{tr}\left(\Sigma^{-1}\frac{\partial \Sigma}{\partial \eta_i}\Sigma^{-1}\frac{\partial^2 \Sigma}{\partial \eta_i^2}\right) + 2\mathrm{tr}\left(\left(\Sigma^{-1}\frac{\partial \Sigma}{\partial \eta_i}\right)^3\right)$$

and similarly for mixed partials, calculations similar to those above show that these expected values are also $O(n)$. Thus the cubic term in the Taylor expansion of the log likelihood is of order $n\|\eta - \hat{\eta}\|^3$. Thus in the relevant range $\|\eta - \hat{\eta}\| = O(n^{-1/2})$, the cubic correction is $O(n^{-1/2})$ and is negligible. Thus Laplace's method applies and we get

$$\begin{aligned}\pi(M|\mathbf{Y}) = {} & \frac{\alpha_M}{Z(\mathbf{Y})(2\pi)^{n/2}}(\det \hat{\Sigma})^{-1/2}\exp(-\frac{1}{2}(\mathbf{Y} - X\hat{\theta}_m)'\hat{\Sigma}^{-1}(\mathbf{Y} - X\hat{\theta}_m))\times \\ & \pi_m(\hat{\theta}_m)\pi_h(\hat{\eta})(2\pi)^{(m+3+2h)/2}(\det I_{\eta,\eta})^{-1/2}(\det I_{\theta,\theta})^{-1/2}\left(1 + O_p(1/n)\right) \\ = {} & \frac{\alpha_M 3^{1/2}}{Z(\mathbf{Y})(2\pi)^{(n-m-3)/2}}(\det \hat{\Sigma})^{-1/2}\exp(-\frac{1}{2}(\mathbf{Y} - X\hat{\theta}_m)'\hat{\Sigma}^{-1}(\mathbf{Y} - X\hat{\theta}_m))\times \\ & \pi_m(\hat{\theta}_m)\pi_h(\hat{\eta})\det(L_m)^{-1/2}n^{-(3m+5)/2}\left(1 + O_p(1/n)\right).\end{aligned} \tag{4.23}$$

If the true model is an $h = 0$ model but the model whose posterior probability we are computing is a full $h = 1$ model, then the Laplace approximation breaks down in a number of ways. First as we see from (4.11), since $\sigma_I^2 = 0$, $\tau^2 = 0$ and the ρ pieces of the information matrix are zero. Physically this corresponds to the fact that if $\sigma_I^2 = 0$, then ρ does not affect the likelihood and therefore we are getting zero information about ρ. Thus the ρ part of the integral cannot be approximated using the Laplace method. Less obvious is that the τ^2 part of the integration cannot be done using the Laplace approximation either. Formula (4.16) shows that we will have $\hat{\tau}^2 = O(n^{-2})$ and the dominant range for the integration will be $|\tau^2 - \hat{\tau}^2| = O(n^{-2})$. Thus the dominant range of the integral will reach the boundary and boundary effects will be significant. If this were the only problem, then it could be handled using the results of Pauler, Wakefield, and Kass (1999). However there is a further problem. In this range the coefficient of the cubic term in the Taylor expansion is

$$\begin{aligned} E\left(\frac{\partial^3 \log L(\eta|Y)}{(\partial\tau^2)^3}\right) &= 2\mathrm{tr}\left(A^{-3}\right) + O(n^4) \\ &= 2n^6 \int_0^1 \int_0^1 \int_0^1 g(t_1,t_2)g(t_2,t_3)g(t_3,t_1)dt_1dt_2dt_3 + O(n^4) \\ &= O(n^6). \end{aligned}$$

In the dominant range, this means that the cubic term is $O(1)$ and is not negligible. Hence naively applying Laplace's method will not give an accurate approximation to the posterior probability.

To obtain an accurate approximation in this case we must use a little more care. Since the (τ,τ) term of the information scales like n^4 and the (β_0,β_0) and (b,b) terms scale like n, we would expect the off-diagonal (τ,β_0) and (τ,b) terms to scale like $n^{2.5}$. Since these off-diagonal entries actually scale like n^2, we see that the parameters τ and (β_0,b) are asymptotically orthogonal. Thus we can split off the integration over β_0 and b and perform it first. Further the dominant contribution comes from $\tau^2 = O(n^{-2})$ and hence $I_{\rho,\rho} = O(n^{-2})$. Thus not only do we get no information about ρ when $\tau^2 = 0$, but over the entire range of integration we get no information about ρ. Thus we may ignore the dependence of the likelihood on ρ in this range. As a result the ρ integral is almost trivial. Let $\tau^2 = \phi^2/n^2$. If $\mathbf{v}$ is a sequence of vectors which converges to a smooth function $f_{\mathbf{v}}(t)$ as n tends to infinity, then

$$n^2\Sigma\mathbf{v} = n^2(A+B)\mathbf{v} \to -e^{2\beta_0}\frac{d}{dt}\left(e^{2bt}\frac{df_{\mathbf{v}}(t)}{dt}\right) + \phi^2 f_{\mathbf{v}}(t).$$

Thus Σ^{-1} will be asymptotic to $n\tilde{g}(s,t)$ where $\tilde{g}$ is the Green's function for this differential operator with boundary conditions $\tilde{g}(s,0) = \tilde{g}(s,1) = 0$. The differential equation

$$-e^{2\beta_0}\frac{d}{dt}\left(e^{2bt}\frac{dy}{dt}\right) + \phi^2 y = 0$$

has solutions

$$y(t) = e^{-bt} I_1\left(\frac{\phi e^{-\beta_0 - bt}}{b}\right) \text{ and } e^{-bt} K_1\left(\frac{\phi e^{-\beta_0 - bt}}{b}\right),$$

where I_1 and K_1 are modified Bessel functions. Hence the Green's function is given by

$$\begin{aligned}\tilde{g}(s,t) = e^{-2\beta_0} & \frac{e^{-b(s+t)}}{b\left(I_1(\phi e^{-\beta_0}/b)K_1(\phi e^{-\beta_0-b}/b) - K_1(\phi e^{-\beta_0}/b)I_1(\phi e^{-\beta_0-b}/b)\right)} \times \\ & \left(I_1(\phi e^{-\beta_0 - b\max(s,t)}/b)K_1(\phi e^{-\beta_0-b}/b) - K_1(\phi e^{-\beta_0 - b\max(s,t)}/b)I_1(\phi e^{-\beta_0-b}/b)\right) \times \\ & \left(I_1(\phi e^{-\beta_0}/b)K_1(\phi e^{-\beta_0 - b\min(s,t)}/b) - K_1(\phi e^{-\beta_0}/b)I_1(\phi e^{-\beta_0 - b\min(s,t)}/b)\right).\end{aligned}$$

Then as for the case $\sigma_I^2 = 0$, we have

$$(X'\Sigma^{-1}X)_{i,j} \sim n^3 \int_0^1 \int_0^1 s^i \tilde{g}(s,t) t^j \, ds dt = n^3 \tilde{\ell}_{i,j},$$

and we define $\tilde{L}_m = (\tilde{\ell}_{i,j})_{0\le i,j\le m}$. Let $m_h(\tau^2, \beta_0, b)$ be the marginal prior for these three parameters, integrating out ρ. Integrating out the θ, β_0, b, and ρ pieces of the posterior probability and substituting $\tau^2 = \phi^2/n^2$ gives

$$\begin{aligned}\pi(M|\mathbf{Y}) = & \frac{\alpha_M 3^{1/2} \pi_m(\hat{\theta}_m) m_h(\hat{\tau}^2, \hat{\beta}_0, \hat{b})}{Z(\mathbf{Y})(2\pi)^{(n-m-3)/2}} n^{-(3m+9)/2} \times \\ & \int_0^\infty \left(\det(\tilde{L}_m)\det(\Sigma)\right)^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{Y} - X\hat{\theta}_m)'\Sigma^{-1}(\mathbf{Y} - X\hat{\theta}_m)\right) d\phi^2.\end{aligned} \tag{4.24}$$

Here in the integral all parameters other than ϕ^2 are to be replaced by their MLEs (ρ should not contribute and may be set to zero). Note that in this case there is no penalty for the parameter ρ but the penalty for σ_I^2 or τ^2 more than compensates. The approximation is further complicated by the fact that the last integral is not Gaussian and cannot be done in closed form.

We now turn to the case where $\sigma_I^2 \neq 0$. In this case the true model is an $h = 1$ model with both the ϵ_j and I_j sources of variation. In this case any $h =$

0 model will be incorrect and therefore will give an exponentially small likelihood and hence an exponentially small posterior probability and an exponentially small BIC. Following Kass and Vaidyanathan (1992) it is not necessary to approximate the posterior probability in this case. Thus we may restrict to the case where the true model is also an $h = 1$ model. Let

$$S = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix}_{n\times n}$$

denote the $n \times n$ cyclic shift matrix. Note that $S^{-1} = S^T = S^{n-1}$ is the cyclic shift in the other direction. Cyclic matrices are matrices which are polynomials in S. Since cyclic matrices commute, they form a convenient subalgebra of all $n \times n$ matrices. If there were no heteroscedasticity in the model, i.e., if $\beta_1 = 0$, then except for negligible boundary effects, the variance components A and B and hence their sum Σ are cyclic matrices. Explicitly

$$\begin{aligned} B &\approx \sigma_I^2 \left(I + \rho(S + S^{-1}) + \rho^2(S^2 + S^{-2}) + \cdots\right) \\ &= (1-\rho^2)\sigma_I^2 \left[(1+\rho^2)I - \rho(S + S^{-1})\right]^{-1}, \quad \text{and} \\ A &\approx e^{2\beta_0}(2I - S - S^{-1}). \end{aligned}$$

If $\beta_1 \neq 0$, then A is not quite so simple. However even in this case A is still a tridiagonal matrix and can be related to cyclic matrices. Let $D = (d_{i,j})$ be the diagonal $n \times n$ matrix with diagonal entries $d_{i,i} = \exp(2i\beta_1) = \exp(2(i/n)b)$. The matrices S and D do not commute, but since the entries of D are slowly varying we have $S^k D \approx DS^k$ for $|k| << n$. Since A, B and Σ are all concentrated near the diagonal only small powers of S will contribute in the formulas below and this will suffice. Thus we may carry out our calculations as though D and S commute. With this definition we have

$$A \approx e^{2\beta_0} D \left(2I - S - S^{-1}\right),$$

and

$$\Sigma \approx e^{2\beta_0} D \left(2I - S - S^{-1}\right) + (1-\rho^2)\sigma_I^2 \left[(1+\rho^2)I - \rho(S + S^{-1})\right]^{-1}. \tag{4.25}$$

Thus

$$\begin{aligned}
\Sigma^{-1} &\approx \left((1+\rho^2)I - \rho(S+S^{-1})\right) \times \\
&\quad \left[(1-\rho^2)\sigma_I^2 I + 2(1+\rho^2)e^{2\beta_0}D - (1+\rho)^2 e^{2\beta_0} D(S+S^{-1}) + \rho e^{2\beta_0} D(S+S^{-1})^2\right]^{-1} \\
&= \left((1+\rho^2)I - \rho(S+S^{-1})\right) \times \\
&\qquad \left[((1-\rho^2)\sigma_I^2 I + 2(1+\rho^2)e^{2\beta_0}D)(I - R_+(S+S^{-1}))(I - R_-(S+S^{-1}))\right]^{-1},
\end{aligned}$$

where $R_\pm$ are the diagonal matrices given by

$$\begin{aligned}
R_\pm = &\left((1+\rho)^2 e^{2\beta_0} D \pm \sqrt{(1-\rho)^4 e^{4\beta_0} D^2 - 4\rho(1-\rho^2)\sigma_I^2 e^{2\beta_0} D}\right) \times \\
&\left[2((1-\rho^2)\sigma_I^2 I + 2(1+\rho^2)e^{2\beta_0}D)\right]^{-1}.
\end{aligned}$$

Let $C_\pm$ be the diagonal matrices given by

$$C_\pm = \left((1+\rho^2)R_\pm - \rho I\right)\left[R_\pm - R_\mp\right]^{-1}.$$

Then we compute

$$
\begin{aligned}
\Sigma^{-1} &\approx \left((1-\rho^2)\sigma_I^2 I + 2(1+\rho^2)e^{2\beta_0}D\right)^{-1} \times \\
&\qquad \left(C_+\left[I - R_+(S+S^{-1})\right]^{-1} + C_-\left[I - R_-(S+S^{-1})\right]^{-1}\right) \\
&= \left((1-\rho^2)\sigma_I^2 I + 2(1+\rho^2)e^{2\beta_0}D\right)^{-1} \sum_{m=0}^{\infty} \left(C_+R_+^m + C_-R_-^m\right)(S+S^{-1})^m \\
&= \left((1-\rho^2)\sigma_I^2 I + 2(1+\rho^2)e^{2\beta_0}D\right)^{-1} \sum_{m=0}^{\infty} \left(C_+R_+^m + C_-R_-^m\right) \sum_{r=0}^{m} \begin{pmatrix} m \\ r \end{pmatrix} S^{m-2r} \\
&= \left((1-\rho^2)\sigma_I^2 I + 2(1+\rho^2)e^{2\beta_0}D\right)^{-1} \times \\
&\qquad \sum_{k=-\infty}^{\infty} \sum_{r=0}^{\infty} \left(C_+R_+^{|k|+2r} + C_-R_-^{|k|+2r}\right) \begin{pmatrix} |k|+2r \\ r \end{pmatrix} S^k \\
&= \left((1-\rho^2)\sigma_I^2 I + 2(1+\rho^2)e^{2\beta_0}D\right)^{-1} \times \\
&\qquad \sum_{k=-\infty}^{\infty} \left\{ C_+\left[I - 4R_+^2\right]^{-1/2} \left(2R_+\left[I + \sqrt{I - 4R_+^2}\right]^{-1}\right)^{|k|} + \right. \\
&\qquad\qquad \left. C_-\left[I - 4R_-^2\right]^{-1/2} \left(2R_-\left[I + \sqrt{I - 4R_-^2}\right]^{-1}\right)^{|k|} \right\} S^k.
\end{aligned}
\tag{4.26}
$$

Dropping the $\pm$ subscripts, the diagonal entries $r_{i,i}$ of R_+ and R_- are the two roots of the quadratic equation

$$
\left((1-\rho^2)\sigma_I^2 + 2(1+\rho^2)e^{2\beta_0}d_{i,i}\right) r_{i,i}^2 - (1+\rho)^2 e^{2\beta_0} d_{i,i} r_{i,i} + \rho e^{2\beta_0} d_{i,i} = 0. \tag{4.27}
$$

If the roots of this polynomial are complex conjugates, then their squared norm (and their product) is

$$
|r_{i,i}|^2 = \frac{\rho e^{2\beta_0} d_{i,i}}{(1-\rho^2)\sigma_I^2 + 2(1+\rho^2)e^{2\beta_0} d_{i,i}}.
$$

For $\sigma_I^2(1-\rho^2) > 0$, we conclude

$$
|r_{i,i}|^2 < \frac{\rho}{2(1+\rho^2)} < \frac{1}{4}.
$$

If the roots are real, then rearranging (4.27) gives

$$(2r_{i,i}-1)((1+\rho^2)r_{i,i}-\rho) = -\frac{(1-\rho^2)\sigma_I^2 r_{i,i}^2}{e^{2\beta_0}d_{i,i}}.$$

Note that for $\sigma_I^2(1-\rho^2) > 0$, the right hand side of this equation is negative. Therefore, $r_{i,i}$ must lie strictly between the two roots of the quadratic on the left. Since $-\frac{1}{2} < \frac{\rho}{1+\rho^2} < \frac{1}{2}$, we conclude that $-\frac{1}{2} < r_{i,i} < \frac{1}{2}$. Combining these two cases, we see that every entry of the diagonal matrices $R_\pm$ has magnitude strictly less than $\frac{1}{2}$. Therefore, the series in (4.26) all converge and the coefficients of S^k decay exponentially as $|k|$ tends to ∞. This justifies our claim above that Σ^{-1} is concentrated near the diagonal and hence our use of the approximation $S^k D \approx DS^k$ was legitimate. Since the coefficients decay exponentially, the coefficient of $S^n = I$ and powers of higher multiples of n are negligible and we can ignore them below.

In the limit as n tends to infinity, the diagonal matrices D, $R_\pm$, and $C_\pm$ should be interpreted as converging to functions on $[0,1]$. The diagonal matrix D converges to the function $d(t) = \exp(2bt)$. Let $r_\pm(t)$ be the limiting functions for $R_\pm$. Then

$$r_\pm = \frac{(1+\rho)^2 e^{2\beta_0+2bt} \pm \sqrt{(1-\rho)^4 e^{4\beta_0+4bt} - 4\rho(1-\rho^2)\sigma_I^2 e^{2\beta_0+2bt}}}{2((1-\rho^2)\sigma_I^2 + 2(1+\rho^2)e^{2\beta_0+2bt})}.$$

The formulas (4.25) and (4.26) give

$$\Sigma \sim \sum_{k=-\infty}^{\infty} F_k S^k \to \sum_{k=-\infty}^{\infty} f_k(t) S^k, \quad \text{and}$$

$$\Sigma^{-1} \sim \sum_{k=-\infty}^{\infty} G_k S^k \to \sum_{k=-\infty}^{\infty} g_k(t) S^k$$

for diagonal matrices F_k and G_k and explicit but complicated limiting functions f_k and g_k. Explicitly

$$f_k(t) = \begin{cases} \sigma_I^2 + 2e^{2\beta_0+2bt} & k=0 \\ \rho\sigma_I^2 - e^{2\beta_0+2bt} & |k|=1 \\ \rho^{|k|}\sigma_I^2 & |k|>1 \end{cases}$$

$$g_k(t) = \frac{1}{\sqrt{(1-\rho)^4 e^{4\beta_0+4bt} - 4\rho(1-\rho^2)\sigma_I^2 e^{2\beta_0+2bt}}} \times$$
$$\left\{ \frac{((1+\rho^2)r_+(t)-\rho)}{\sqrt{1-4r_+^2(t)}} \left(\frac{2r_+(t)}{1+\sqrt{1-4r_+^2(t)}} \right)^{|k|} + \right.$$
$$\left. \frac{((1+\rho^2)r_-(t)-\rho)}{\sqrt{1-4r_-^2(t)}} \left(\frac{2r_-(t)}{1+\sqrt{1-4r_-^2(t)}} \right)^{|k|} \right\}.$$

Note that $f_{-k} = f_k$ and $g_{-k} = g_k$. Therefore

$$\begin{aligned} I_{\eta_i,\eta_j} &= -\frac{1}{2}\mathrm{tr}\left(\frac{\partial\Sigma^{-1}}{\partial\eta_i}\frac{\partial\Sigma}{\partial\eta_j}\right) \\ &\sim -\frac{1}{2}\mathrm{tr}\left(\sum_{k=-\infty}^{\infty}\frac{\partial G_k}{\partial\eta_i}S^k\sum_{\ell=-\infty}^{\infty}\frac{\partial F_\ell}{\partial\eta_j}S^\ell\right) \\ &\sim -\frac{1}{2}\sum_{k=-\infty}^{\infty}\sum_{\ell=-\infty}^{\infty}\mathrm{tr}\left(\frac{\partial G_k}{\partial\eta_i}\left[S^k\frac{\partial F_\ell}{\partial\eta_j}S^{-k}\right]S^{k+\ell}\right). \end{aligned}$$

We will see below that this sum converges exponentially, therefore we need only consider terms with $|k|, |\ell| << n$. Hence S^k and $\frac{\partial F_\ell}{\partial \eta_j}$ approximately commute. Since S is a cyclic shift matrix, $S^{k+\ell}$ has only zero entries on the diagonal unless $k+\ell$ is a multiple of n. Since $|k|, |\ell| << n$, the only case we need consider is when $k+\ell = 0$. Plugging in these two observations gives

$$\begin{aligned} I_{\eta_i,\eta_j} &\sim -\frac{1}{2}\sum_{k=-\infty}^{\infty}\mathrm{tr}\left(\frac{\partial G_k}{\partial\eta_i}\frac{\partial F_{-k}}{\partial\eta_j}\right) \\ &\sim -\frac{n}{2}\sum_{k=-\infty}^{\infty}\int_0^1\left(\frac{\partial g_k(t)}{\partial\eta_i}\frac{\partial f_k(t)}{\partial\eta_j}\right)dt. \end{aligned}$$

The functions g_k and f_k decay exponentially as $|k|$ tends to infinity. Hence this sum converges rapidly and we see that

$$I_{\eta,\eta} = nK(\eta) + O(1)$$

for some calculable 4×4 matrix $K(\eta)$. In particular all eigenvalues of the information matrix tend to infinity as n tends to infinity. Thus the integral representing the posterior probability of model M is peaked and the dominant contribution comes from η with $\|\eta - \hat{\eta}\| = O(n^{-1/2})$. Similar arguments show that the cubic and higher order coefficients in the Taylor expansion of the log likelihood are also $O(n)$. Thus they are negligible for η in the dominant range. Hence the standard Laplace approximation applies in this case and we obtain

$$\begin{aligned}\pi(M|\mathbf{Y}) = \frac{\alpha_M \pi_m(\hat{\theta}_m)\pi_h(\hat{\eta})}{Z(\mathbf{Y})(2\pi)^{(n-m-5)/2}}&(\det\hat{\Sigma})^{-1/2}\exp(-\frac{1}{2}(\mathbf{Y}-X\hat{\theta}_m)'\hat{\Sigma}^{-1}(\mathbf{Y}-X\hat{\theta}_m))\times \\ &n^{-(m+5)/2}(\det K(\hat{\eta}))^{-1/2}\left(\frac{\hat{\sigma}_I^2(1+\hat{\rho})}{1-\hat{\rho}}\right)^{(m+1)/2}\prod_{i=1}^{m}(2i+1)^{1/2}\binom{2i}{i}\times \\ &(1+O_p(1/n)).\end{aligned} \tag{4.28}$$

This formula, which is valid for $\hat{\sigma}_I^2 = O(1)$, and the previous formula, which is valid for $\hat{\sigma}_I^2 = O(n^{-2})$, cover the two cases possible under our model, but do not cover the intermediate range. It would be nice to have an approximation which interpolates between the two cases, even if the asymptotic behavior of this interpolating formula is not clear. To derive a candidate for this formula look at the posterior probability integral for an $h = 1$ model

$$\begin{aligned}\pi(M|\mathbf{Y}) &= \frac{\alpha_M}{Z(\mathbf{Y})}\int_{\Omega_1}\int_{\Theta_m}\pi_1(\eta)\pi_m(\theta_m)L(\theta_m,\eta|\mathbf{Y})d\theta_m d\eta \\ &= \frac{\alpha_M}{(2\pi)^{n/2}Z(\mathbf{Y})}\int_{\Omega_1}\int_{\Theta_m}\frac{\pi_1(\eta)\pi_m(\theta_m)}{\det(\Sigma)^{1/2}}\exp\left(-\frac{1}{2}(\mathbf{Y}-X\theta_m)'\Sigma^{-1}(\mathbf{Y}-X\theta_m)\right)d\theta_m d\eta.\end{aligned}$$

Let $\tilde{\theta}_m = \tilde{\theta}_m(\eta)$ denote the maximum likelihood estimator of θ_m given the value of η. Since all eigenvalues of $I_{\theta,\theta} = X'\Sigma^{-1}X$ tend to infinity as n tends to infinity regardless of η, the inner integral over θ_m is sharply peaked near $\tilde{\theta}_m$. Thus the error in replacing $\pi_m(\theta_m)$ by $\pi_m(\tilde{\theta}_m)$ in the inner integral is small. After making this replacement the inner integral can be done exactly, yielding

$$
\begin{aligned}
\pi(M|\mathbf{Y}) \approx{}& \frac{\alpha_M}{(2\pi)^{n/2} Z(\mathbf{Y})} \int_{\Omega_1} \frac{\pi_1(\eta)\pi_m(\tilde{\theta}_m)}{\det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{Y} - X\tilde{\theta}_m)'\Sigma^{-1}(\mathbf{Y} - X\tilde{\theta}_m)\right) \times \\
& \int_{\Theta_m} \exp\left(-\frac{1}{2}(\theta_m - \tilde{\theta}_m)' X' \Sigma^{-1} X (\theta_m - \tilde{\theta}_m)\right) d\theta_m d\eta \\
={}& \frac{\alpha_M}{(2\pi)^{(n-m-1)/2} Z(\mathbf{Y})} \int_{\Omega_1} \frac{\pi_1(\eta)\pi_m(\tilde{\theta}_m)}{\det(\Sigma)^{1/2}} \det(X'\Sigma^{-1}X)^{-1/2} \times \\
& \exp\left(-\frac{1}{2}(\mathbf{Y} - X\tilde{\theta}_m)'\Sigma^{-1}(\mathbf{Y} - X\tilde{\theta}_m)\right) d\eta \\
={}& \frac{\alpha_M (2\pi)^{(m+1)/2}}{Z(\mathbf{Y})} \int_{\Omega_1} \pi_1(\eta)\pi_m(\tilde{\theta}_m) \det(X'\Sigma^{-1}X)^{-1/2} L(\tilde{\theta}_m, \eta|\mathbf{Y}) d\eta.
\end{aligned}
$$

Essentially we have done the Laplace approximation over the parameters θ_m, but not over η. If σ_I^2 is in the intermediate range, $O(n^{-2}) < \sigma_I^2 = o(1)$, then we would still expect the B part of Σ to dominate the A part in $X'\Sigma^{-1}X$. If this is the case, then $\det(X'\Sigma^{-1}X)^{-1/2}$ will be proportional to $(\sigma_I^2(1+\rho)/(1-\rho))^{-(m+1)/2}$. Thus this determinant will be a rapidly varying function. Hence it is not valid to apply Laplace's method with this determinant included in the slowly varying term $b(\eta)$ in (2.1). Instead we can include it in the rapidly varying exponential piece (see (2.1)) by setting

$$
h(\eta, n) = \log L(\tilde{\theta}_m, \eta|\mathbf{Y}) - \frac{1}{2} \log\det(X'\Sigma^{-1}X).
$$

Let $\tilde{\eta}$ be the value that maximizes $h(\eta, n)$, then

$$
\pi(M|\mathbf{Y}) \approx \frac{\alpha_M (2\pi)^{(m+5)/2}}{Z(\mathbf{Y})} \pi_1(\tilde{\eta})\pi_m(\tilde{\theta}_m) \det(X'\tilde{\Sigma}^{-1}X)^{-1/2} \det(H_n)^{-1/2} L(\tilde{\theta}_m, \tilde{\eta}|\mathbf{Y}), \tag{4.29}
$$

where $H_n = -\frac{\partial^2}{\partial\eta^2} h(\eta, n)|_{\tilde{\eta}}$ is the Hessian. The n dependence of this formula is not as explicit as (4.24) or (4.28) but it should be valid for intermediate values of σ_I^2.

CHAPTER V

NUMERICAL RESULTS

5.1 Analysis of Selected Variable Stars

Having completed the asymptotic analysis of the posterior probability in the previous chapters, we would like to understand how well the approximations work for actual data sets, each with a finite sample size, n. The first step in this comparison is to choose explicit priors. First consider the prior probabilities $\alpha_{(m,h)}$ on the models. There seems no *a priori* reason to prefer one covariance structure over the other. Therefore, we took the priors to be independent of h. The main interest in the Star Model is to test for whether or not there is a trend in the Y_js. Therefore, we assigned a combined prior of 1/2 (or 1/4 each) to the two no-trend ($m = 0$) models. For the remaining degrees, we chose a prior proportional to $1/m$. Normalizing these accordingly gives

$$\alpha_{(m,h)} = \begin{cases} 1/4 & \text{if } m = 0 \\ 0.0753413946/m & \text{if } m = 1, \ldots, 15. \end{cases}$$

For the priors on the model parameters, we assumed the mean parameters θ_m were a priori independent of the covariance parameters η. We elected to use multivariate normal priors on the θ_m, i.e., $\theta_m \sim N_{m+1}(\nu_m, W_m)$, where ν_m and W_m were obtained from data as will be discussed below. The choice of a multivariate prior for θ_m was in part motivated by the fact that it allowed the computation of the posterior probability integral over θ_m to be done in closed form. The covariance parameters β_0 and b come from measurement error, whereas the parameters ρ and σ_Z^2 come from intrinsic variation within the star. Therefore we assumed that (β_0, b) and (ρ, σ_Z^2) were *a*

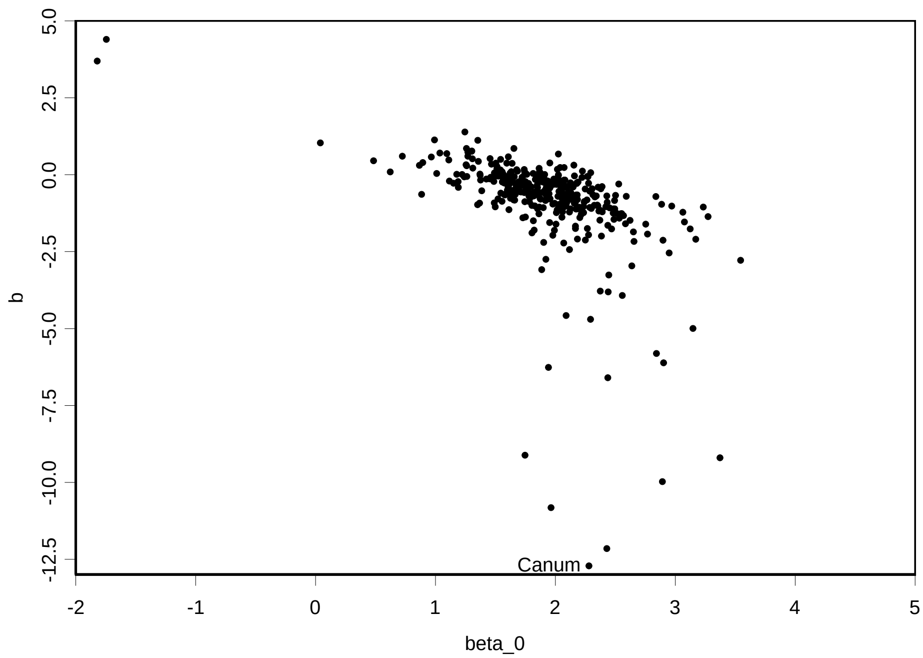

Figure 1. Scatterplot of $\hat{b}$ vs. $\hat{\beta}_0$

priori independent. For (β_0, b) we chose a bivariate normal prior, $(\beta_0, b) \sim N_2(\gamma, V)$. A scatterplot of the observed $(\hat{\beta}_0, \hat{b})$ pairs for the data is shown in Figure 1.

Since ρ is confined to the range $[-1, 1]$, we took it to be *a priori* uniformly distributed on $[-1, 1]$. In the AR(1) model, we have

$$I_j = \rho I_{j-1} + Z_j, \quad j = 2, \ldots, n$$

where $Z_2, \ldots, Z_n$ are iid $N(0, \sigma_Z^2)$. Here ρ represents the carry-over from the previous observation and Z_j a new random effect. Therefore we chose to have ρ and σ_Z^2 be *a priori* independent. Since $\sigma_Z^2 \geq 0$, we chose to use a Gamma distribution for σ_Z^2. We further wanted to keep the prior bounded and nonzero as $\sigma_Z^2 \to 0$, and therefore were compelled to use an exponential prior for σ_Z^2. A histogram of the observed $\hat{\sigma}_Z^2$ for the data and the fitted exponential density are shown in Figure 2. The agreement

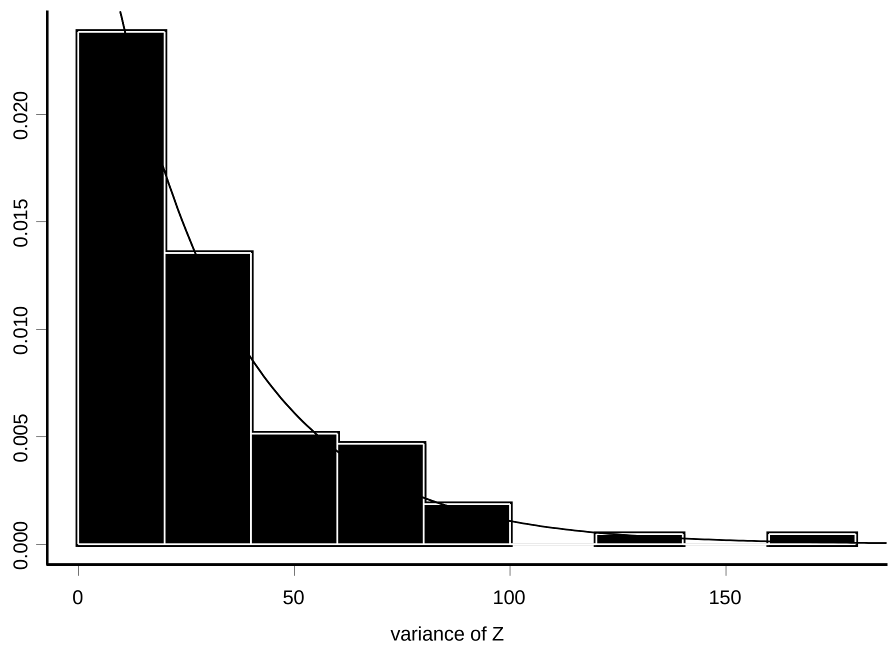

Figure 2. Histogram of observed $\hat{\sigma}_Z^2$ and fitted density

between the fit and the data is reasonably good.

In order to choose the means and covariance matrices for the multivariate normal priors and the mean of the exponential prior on σ_Z^2, we used data from 378 variable stars. For each star we chose a putative correct model $(\hat{m}_i, \hat{h}_i)$ using a naive application of BIC (which does not require specifying a prior) and computed the maximum likelihood estimators of the parameters for the selected model. The prior mean ν_m and prior covariance W_m for θ_m were chosen to be the sample mean and sample covariance of $\hat{\theta}_m$ for all stars having $\hat{m} = m$. Two exceptions to this rule were required. The number of stars with $\hat{m} = 14$ and $\hat{m} = 15$ was too small to give a positive definite sample covariance W_m. Therefore data for $\hat{m} = 14$ and $\hat{m} = 15$ were pooled to give estimates for W_{14} and W_{15}. Also one star with only 32 observations had $\hat{m} = 15$. The theory requires the number of observations to be much larger than the degree, which was certainly false for this star and the resulting $\hat{\theta}_m$ was unreasonable. Therefore this star was excluded from the computations. Similarly the prior mean γ and prior covariance V for (β_0, b) were taken to be the sample mean and sample covariance of $(\hat{\beta}_0, \hat{b})$ for all stars, excluding three outliers. The estimators $\hat{\sigma}_Z^2$ roughly followed an exponential distribution and we chose the prior mean to be the sample mean of $\hat{\sigma}_Z^2$.

The main step in computing the exact posterior probabilities is to evaluate the integrals

$$J_M = \alpha_M \int_{\Omega_h} \int_{\Theta_m} L(\theta_m, \eta_h | \mathbf{Y}) \pi_m(\theta_m) \pi_h(\eta_h) d\theta_m d\eta.$$

Let $X = X_m$ be the $n \times (m+1)$ design matrix for the m-th degree polynomial model. Then we have

$$J_M = \frac{\alpha_M}{(2\pi)^{(m+n+1)/2} \det(W_m)^{1/2}} \int_{\Omega_h} \frac{\pi_h(\eta_h)}{\det(\Sigma(\eta))^{1/2}} \times$$
$$\int_{\Theta_m} \exp\left(-\frac{1}{2}(\mathbf{Y} - X\theta_m)' \Sigma^{-1} (\mathbf{Y} - X\theta_m) - \frac{1}{2}(\theta_m - \nu_m)' W_m^{-1} (\theta_m - \nu_m) \right) d\theta_m d\eta_h.$$

The integrand in the inner integral is an unnormalized multivariate normal density and therefore this integral can be done in closed form, giving

$$J_M = \frac{\alpha_M e^{-\nu_m' W_m^{-1} \nu_m/2}}{(2\pi)^{n/2} \det(W_m)^{1/2}} \int_{\Omega_h} \frac{\pi_h(\eta_h)}{\det(\Sigma(\eta))^{1/2} \det(W_m^{-1} + X'\Sigma^{-1}X)^{1/2}} \times$$
$$\exp\Big(-\frac{1}{2}\mathbf{Y}'\Sigma^{-1}\mathbf{Y} + \frac{1}{2}(\mathbf{Y}'\Sigma^{-1}X + \nu_m' W_m^{-1}) \left[X'\Sigma^{-1}X + W_m^{-1} \right]^{-1} \times$$
$$(X'\Sigma^{-1}\mathbf{Y} + W_m^{-1}\nu_m) \Big) d\eta_h.$$

This leaves either a 2-dimensional or a 4-dimensional integral to be done numerically. These numerical integrals were done using importance sampling by simulating 10,000 observations from a multivariate normal distribution with the same mode and Hessian at the mode as the integrand.

The first formula we would like to compare this to is the standard BIC approximation. Specifically, we define

$$\mathrm{BIC}_M = 2 \log L_M(\hat{\theta}_m, \hat{\eta} | \mathbf{Y}) - (m + 3 + 2h) \log n,$$

where n is the sample size and $m+3+2h$ is the total number of parameters in model M. Then the resulting approximation to the posterior probability of model M is

$$\pi(M|\mathbf{Y}) \approx \frac{e^{\mathrm{BIC}_M/2}}{\sum_{M'} e^{\mathrm{BIC}_{M'}/2}}.$$

The second approximation we would like to consider is a modified version of BIC which has the correct asymptotic behavior for our models. Specifically, define

$$\mathrm{mod}-\mathrm{BIC}_M = 2\log L_M(\hat{\theta}_m, \hat{\eta}|\mathbf{Y}) - k_M \log n,$$

where the penalty k_M for model M is given by

$$k_M = \begin{cases} 3m+5 & \text{if } h=0 \\ m+5 & \text{if } h=1 \text{ and } \hat{\sigma}_Z^2 \geq 0.001 \\ 3m+9 & \text{if } h=1 \text{ and } \hat{\sigma}_Z^2 < 0.001 \end{cases}$$

These penalties have been chosen based on the powers of n in (4.23), (4.24), and (4.28). The cutoff $\hat{\sigma}_Z^2 < 0.001$ for declaring that the true model is the $h=0$ model and hence (4.24) applies is somewhat arbitrary. However, the data examples with $\hat{\sigma}_Z^2 < 0.001$ invariably had much smaller $\hat{\sigma}_Z^2$.

The third approximation is the basic Laplace approximation (4.1), where the information matrices $I_{\theta,\theta}$ and $I_{\eta,\eta}$ are estimated as the negative of the Hessian of the log-likelihood at the MLE. The fourth approximation uses (4.1) for $h = 0$ models, but uses the modified version of (4.29) for the $h = 1$ models. The fifth approximation is the asymptotically correct version of the Laplace approximation using (4.23) for $h = 0$ models, (4.24) for $h = 1$ models with $\hat{\sigma}_Z^2 < 0.001$, and (4.28) for $h = 1$ models with $\hat{\sigma}_Z^2 \geq 0.001$.

The "exact" posterior probabilities and the five approximations were computed for 6 stars: Mira, R. Aquilae, R. Bootis, R. Canum Venticorum, W. Aquarii and X. Draconis. The first four of these variable stars were previously analyzed in Hart, Koen and Lombard (2004). The "exact" posterior probabilities of the models and the five approximates are given in Tables 1-6. Plots of the posterior probabilities of each polynomial degree are given in Figures 3-8. Plots of the observed pseudo-periods and the polynomial fits are shown in Figures 9-14.

Table 1. Posterior probabilities for Mira

m	h	"exact"	Std. BIC	Mod. BIC	Std. Lapl.	Mod. Lapl.	Asy. Lapl.
0	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0	0.0042	0.0016	0.0000	0.0039	0.0039	0.0026
5	0	0.2445	0.3160	0.0000	0.2563	0.2549	0.1981
6	0	0.2577	0.4831	0.0000	0.2864	0.2850	0.2368
7	0	0.1373	0.0586	0.0000	0.1450	0.1442	0.1254
8	0	0.2487	0.0936	0.0000	0.2037	0.2026	0.1961
9	0	0.0173	0.0297	0.0000	0.0025	0.0025	0.0026
10	0	0.0340	0.0036	0.0000	0.0016	0.0015	0.0020
11	0	0.0077	0.0007	0.0000	0.0041	0.0041	0.0011
12	0	0.0093	0.0001	0.0000	0.0077	0.0076	0.0000
13	0	0.0004	0.0000	0.0000	0.0000	0.0000	0.0000
14	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
15	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0	1	0.0145	0.0000	0.6063	0.0784	0.0868	0.0580
1	1	0.0000	0.0000	0.3486	0.0000	0.0000	0.0000
2	1	0.0014	0.0000	0.0401	0.0079	0.0009	0.0072
3	1	0.0008	0.0000	0.0049	0.0027	0.0003	0.0025
4	1	0.0078	0.0000	0.0000	0.0000	0.0007	0.0085
5	1	0.0024	0.0042	0.0000	0.0000	0.0001	0.0458
6	1	0.0034	0.0064	0.0000	0.0000	0.0001	0.0507
7	1	0.0018	0.0008	0.0000	0.0000	0.0000	0.0321
8	1	0.0046	0.0012	0.0000	0.0000	0.0042	0.0280
9	1	0.0005	0.0004	0.0000	0.0000	0.0000	0.0003
10	1	0.0012	0.0000	0.0000	0.0000	0.0000	0.0002
11	1	0.0002	0.0000	0.0000	0.0000	0.0001	0.0006
12	1	0.0003	0.0000	0.0000	0.0000	0.0003	0.0014
13	1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
14	1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
15	1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 2. Posterior probabilities for R. Aquilae

m	h	"exact"	Std. BIC	Mod. BIC	Std. Lapl.	Mod. Lapl.	Asy. Lapl.
0	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
5	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
6	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
7	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
8	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
9	0	0.0000	0.0020	0.0000	0.0000	0.0000	0.0000
10	0	0.0000	0.3630	0.0000	0.0000	0.0000	0.0000
11	0	0.9770	0.5088	0.0000	0.9815	0.7700	0.5910
12	0	0.0000	0.0880	0.0000	0.0000	0.0000	0.0000
13	0	0.0000	0.0228	0.0000	0.0000	0.0000	0.0000
14	0	0.0000	0.0026	0.0000	0.0000	0.0000	0.0000
15	0	0.0000	0.0012	0.0000	0.0000	0.0000	0.0000
0	1	0.0016	0.0000	0.0002	0.0174	0.0000	0.0614
1	1	0.0007	0.0000	0.1373	0.0000	0.0000	0.0000
2	1	0.0003	0.0000	0.0189	0.0000	0.0000	0.0000
3	1	0.0003	0.0001	0.7505	0.0000	0.0000	0.0000
4	1	0.0001	0.0000	0.0828	0.0000	0.0000	0.0000
5	1	0.0001	0.0000	0.0091	0.0000	0.0000	0.0000
6	1	0.0000	0.0000	0.0010	0.0000	0.0000	0.0000
7	1	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000
8	1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
9	1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
10	1	0.0000	0.0042	0.0000	0.0000	0.0000	0.0000
11	1	0.0199	0.0059	0.0000	0.0012	0.2300	0.3476
12	1	0.0000	0.0010	0.0000	0.0000	0.0000	0.0000
13	1	0.0000	0.0003	0.0000	0.0000	0.0000	0.0000
14	1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
15	1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0043

Table 3. Posterior probabilities for R. Bootis

m	h	"exact"	Std. BIC	Mod. BIC	Std . Lapl.	Mod. Lapl.	Asy. Lapl.
0	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
5	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
6	0	0.0000	0.0006	0.0000	0.0000	0.0000	0.0000
7	0	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000
8	0	0.0000	0.0104	0.0000	0.0000	0.0000	0.0000
9	0	0.0000	0.1297	0.0000	0.0000	0.0000	0.0000
10	0	0.0000	0.1819	0.0000	0.0000	0.0000	0.0000
11	0	0.0000	0.1403	0.0000	0.0000	0.0000	0.0000
12	0	0.0000	0.0576	0.0000	0.0000	0.0000	0.0000
13	0	0.0000	0.0062	0.0000	0.0000	0.0000	0.0000
14	0	0.0000	0.0096	0.0000	0.0000	0.0000	0.0000
15	0	0.0000	0.0015	0.0000	0.0000	0.0000	0.0000
0	1	0.9667	0.4071	0.8898	0.9565	0.8594	0.9552
1	1	0.0291	0.0428	0.0935	0.0381	0.1344	0.0392
2	1	0.0029	0.0061	0.0134	0.0040	0.0052	0.0041
3	1	0.0002	0.0006	0.0014	0.0007	0.0007	0.0007
4	1	0.0001	0.0001	0.0001	0.0006	0.0000	0.0005
5	1	0.0006	0.0004	0.0008	0.0001	0.0002	0.0002
6	1	0.0001	0.0004	0.0009	0.0000	0.0000	0.0000
7	1	0.0001	0.0000	0.0001	0.0000	0.0000	0.0000
8	1	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
9	1	0.0000	0.0011	0.0000	0.0000	0.0000	0.0000
10	1	0.0002	0.0016	0.0000	0.0000	0.0001	0.0000
11	1	0.0000	0.0012	0.0000	0.0000	0.0000	0.0000
12	1	0.0000	0.0005	0.0000	0.0000	0.0000	0.0000
13	1	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000
14	1	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000
15	1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 4. Posterior probabilities for R. Canum Venaticorum

m	h	“exact”	Std. BIC	Mod. BIC	Std . Lapl.	Mod. Lapl.	Asy. Lapl.
0	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
5	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
6	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
7	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
8	0	0.0000	0.0011	0.0000	0.0000	0.0001	0.0000
9	0	0.0000	0.0004	0.0000	0.0000	0.0000	0.0000
10	0	0.0000	0.1193	0.0000	0.0000	0.0000	0.0000
11	0	0.0000	0.0188	0.0000	0.0000	0.0001	0.0000
12	0	0.0000	0.0046	0.0000	0.0000	0.0000	0.0000
13	0	0.0000	0.0176	0.0000	0.0000	0.0000	0.0000
14	0	0.0000	0.0028	0.0000	0.0000	0.0000	0.0000
15	0	0.0000	0.0003	0.0000	0.0000	0.0000	0.0000
0	1	0.5130	0.6449	0.7724	0.0506	0.2705	0.0663
1	1	0.0023	0.1390	0.1665	0.0000	0.0000	0.0000
2	1	0.0873	0.0423	0.0507	0.0838	0.3299	0.1148
3	1	0.0458	0.0048	0.0057	0.0541	0.1993	0.0725
4	1	0.2024	0.0032	0.0038	0.0610	0.1741	0.0621
5	1	0.0607	0.0004	0.0004	0.0076	0.0260	0.0069
6	1	0.0139	0.0003	0.0004	0.1670	0.0000	0.1732
7	1	0.0184	0.0000	0.0000	0.1969	0.0000	0.1956
8	1	0.0241	0.0000	0.0000	0.1954	0.0000	0.1678
9	1	0.0113	0.0000	0.0000	0.1353	0.0000	0.1137
10	1	0.0106	0.0000	0.0000	0.0034	0.0000	0.0021
11	1	0.0045	0.0000	0.0000	0.0449	0.0000	0.0250
12	1	0.0021	0.0000	0.0000	0.0000	0.0000	0.0000
13	1	0.0019	0.0000	0.0000	0.0000	0.0000	0.0000
14	1	0.0009	0.0000	0.0000	0.0000	0.0000	0.0000
15	1	0.0008	0.0000	0.0000	0.0000	0.0000	0.0000

Table 5. Posterior probabilities for W. Aquarii

m	h	"exact"	Std. BIC	Mod. BIC	Std . Lapl.	Mod. Lapl.	Asy. Lapl.
0	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
5	0	0.0000	0.0002	0.0000	0.0000	0.0000	0.0000
6	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
7	0	0.0000	0.0037	0.0000	0.0000	0.0000	0.0000
8	0	0.0000	0.0004	0.0000	0.0000	0.0000	0.0000
9	0	0.0000	0.0026	0.0000	0.0000	0.0000	0.0000
10	0	0.0000	0.0004	0.0000	0.0000	0.0000	0.0000
11	0	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000
12	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
13	0	0.0173	0.4254	0.0000	0.0275	0.0011	0.0000
14	0	0.0000	0.0683	0.0000	0.0000	0.0000	0.0000
15	0	0.0000	0.0416	0.0000	0.0000	0.0000	0.0000
0	1	0.1771	0.3814	0.8490	0.4467	0.0238	0.4096
1	1	0.0001	0.0478	0.1063	0.0000	0.0000	0.0000
2	1	0.0664	0.0087	0.0193	0.2375	0.0140	0.2485
3	1	0.0707	0.0029	0.0065	0.1891	0.0073	0.1966
4	1	0.0398	0.0009	0.0019	0.0230	0.0074	0.0219
5	1	0.1120	0.0066	0.0146	0.0569	0.0079	0.0465
6	1	0.0145	0.0008	0.0018	0.0041	0.0007	0.0030
7	1	0.0384	0.0002	0.0004	0.0037	0.0000	0.0018
8	1	0.0180	0.0000	0.0001	0.0000	0.0000	0.0000
9	1	0.0378	0.0000	0.0000	0.0102	0.0000	0.0037
10	1	0.0143	0.0000	0.0000	0.0000	0.0000	0.0000
11	1	0.0019	0.0000	0.0000	0.0000	0.0000	0.0000
12	1	0.0004	0.0000	0.0000	0.0000	0.0000	0.0000
13	1	0.3895	0.0063	0.0000	0.0013	0.9379	0.0685
14	1	0.0010	0.0010	0.0000	0.0000	0.0000	0.0000
15	1	0.0009	0.0006	0.0000	0.0000	0.0000	0.0000

Table 6. Posterior probabilities for X. Draconis

m	h	"exact"	Std. BIC	Mod. BIC	Std . Lapl.	Mod. Lapl.	Asy. Lapl.
0	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
5	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
6	0	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000
7	0	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000
8	0	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000
9	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
10	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
11	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
12	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
13	0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
14	0	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000
15	0	0.0000	0.0004	0.0000	0.0000	0.0000	0.0000
0	1	0.7655	0.8872	0.8879	0.7417	0.6212	0.7543
1	1	0.2175	0.0945	0.0946	0.2443	0.3395	0.2339
2	1	0.0107	0.0154	0.0155	0.0090	0.0303	0.0079
3	1	0.0034	0.0017	0.0017	0.0032	0.0057	0.0026
4	1	0.0022	0.0002	0.0002	0.0016	0.0027	0.0012
5	1	0.0003	0.0001	0.0001	0.0002	0.0003	0.0001
6	1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
7	1	0.0001	0.0000	0.0000	0.0000	0.0001	0.0000
8	1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
9	1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
10	1	0.0001	0.0000	0.0000	0.0000	0.0002	0.0000
11	1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
12	1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
13	1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
14	1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
15	1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

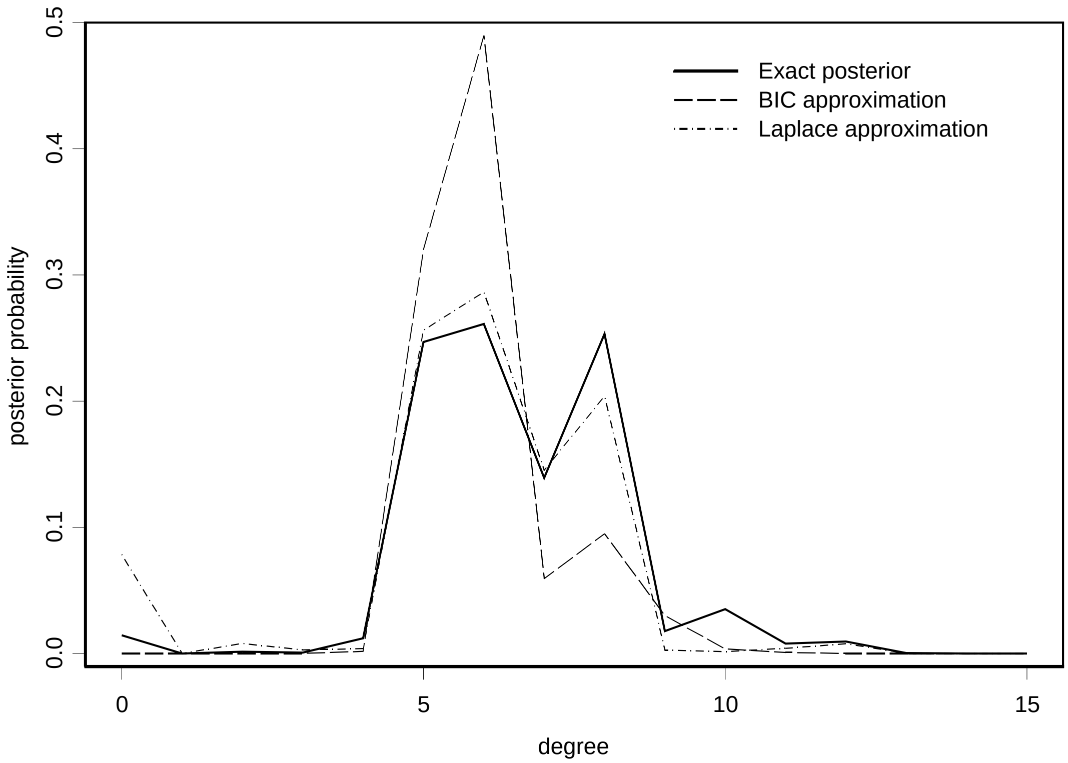

Figure 3. Posterior probabilities for Mira

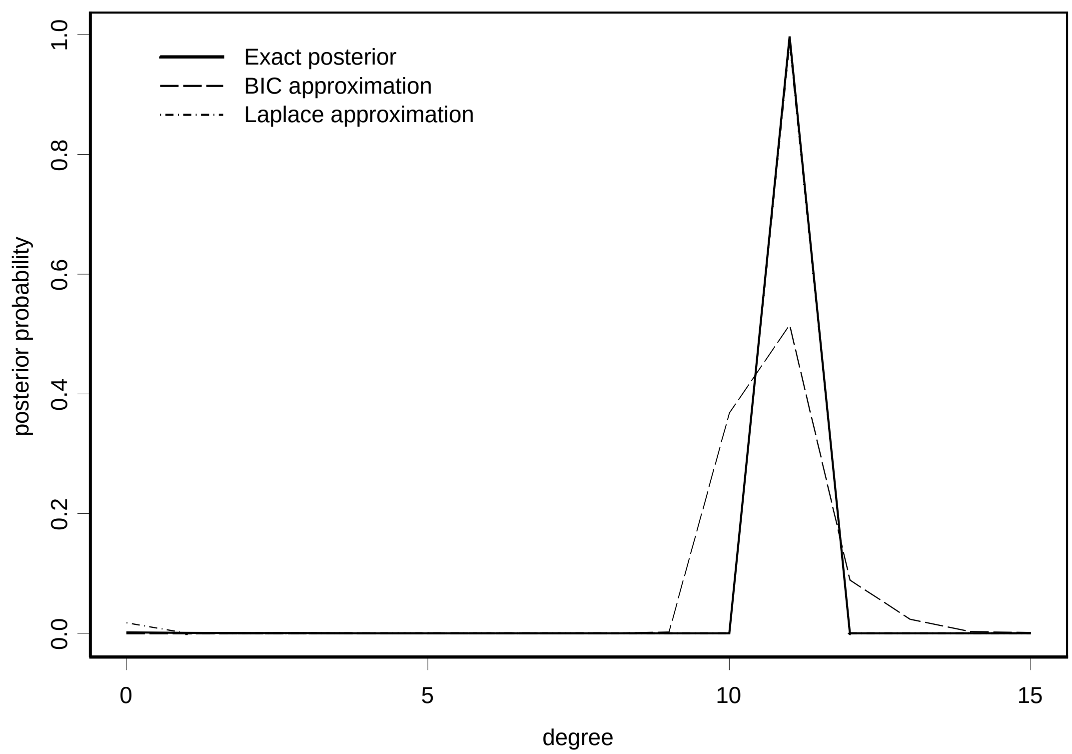

Figure 4. Posterior probabilities for R. Aquilae

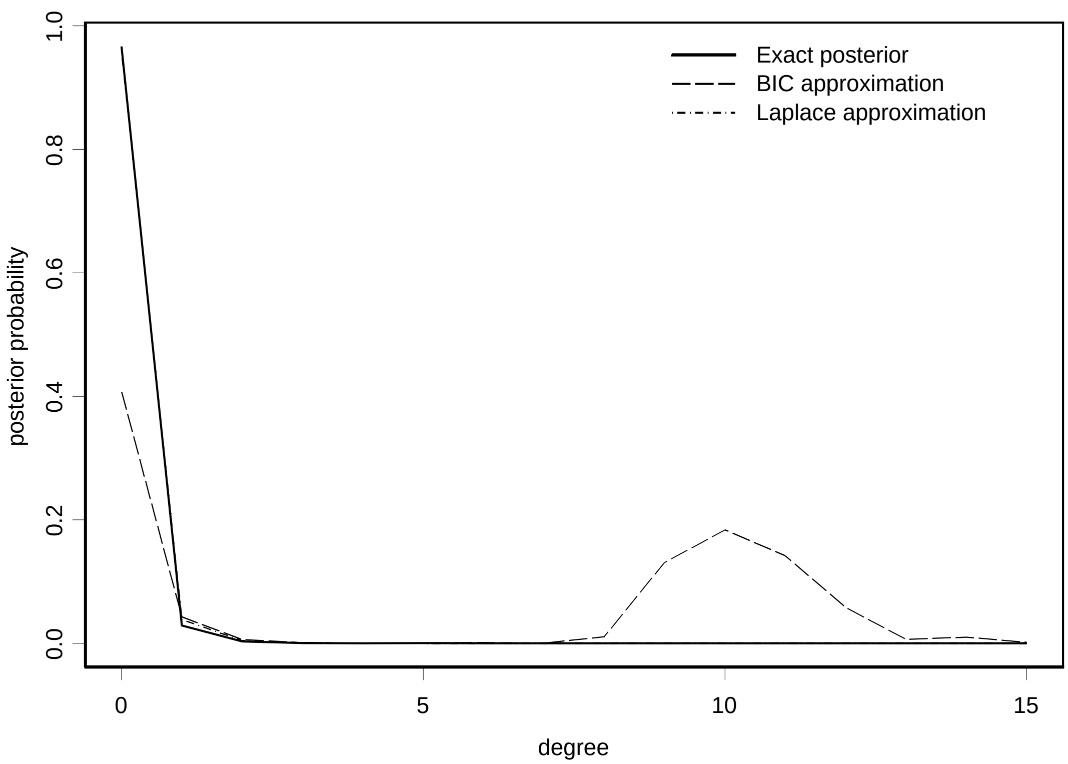

Figure 5. Posterior probabilities for R. Bootis

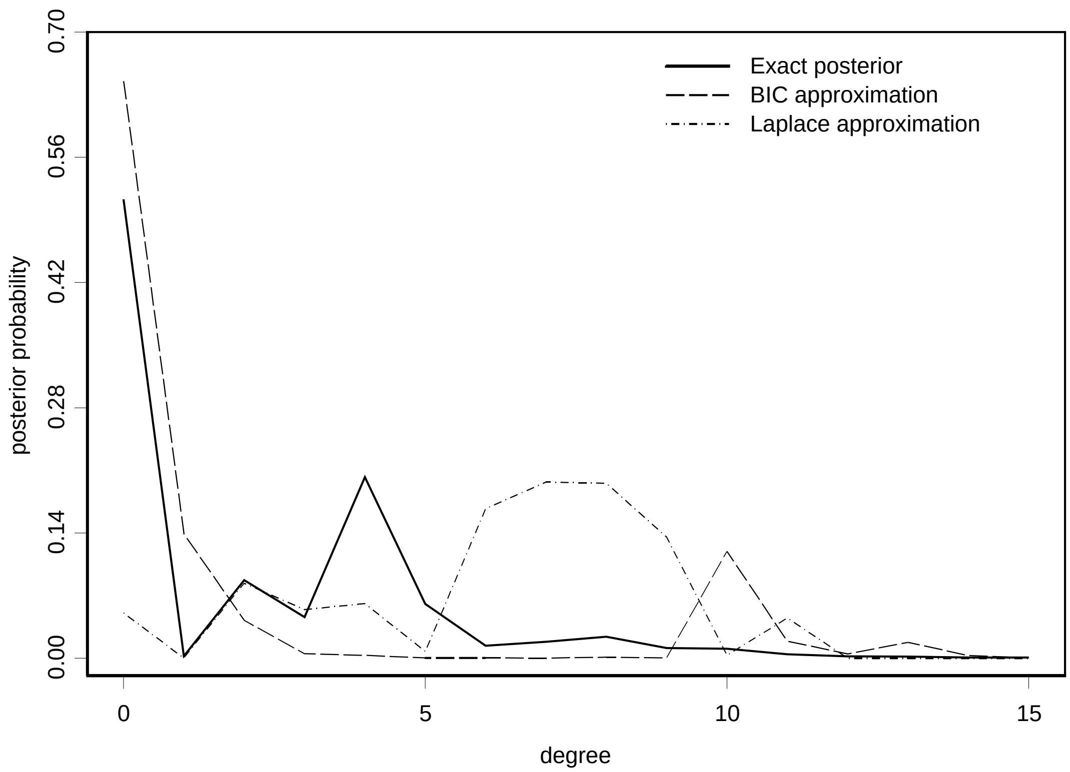

Figure 6. Posterior probabilities for R. Canum Venaticorum

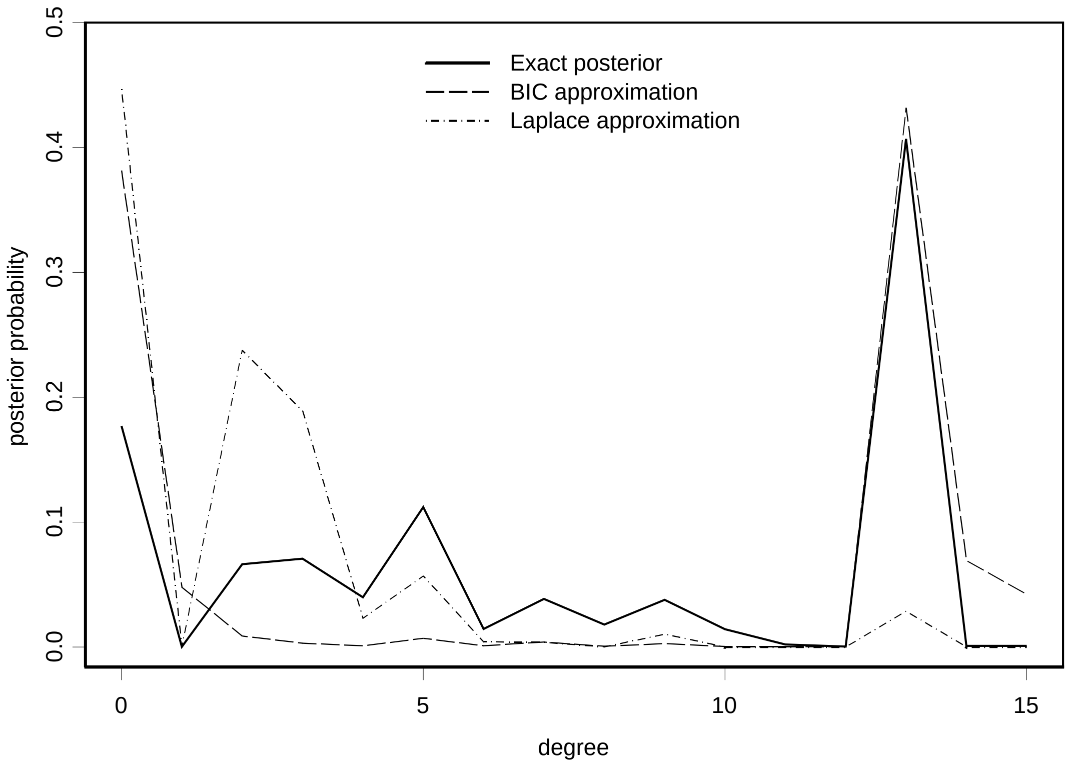

Figure 7. Posterior probabilities for W. Aquarii

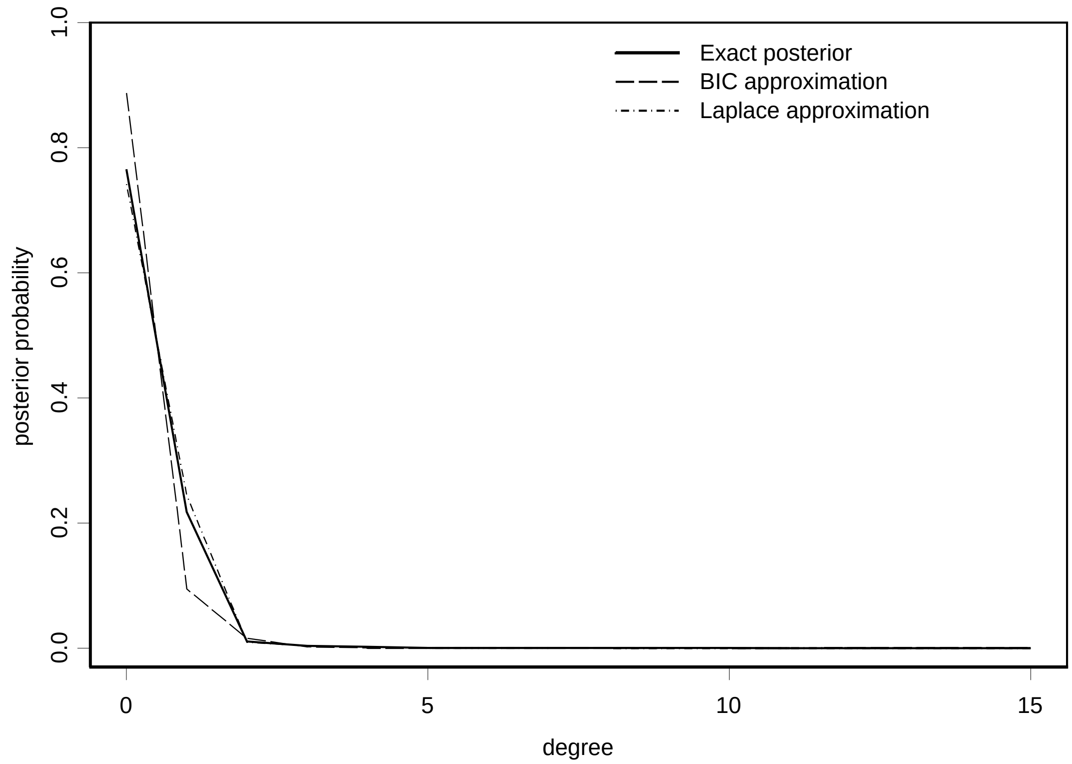

Figure 8. Posterior probabilities for X. Draconis

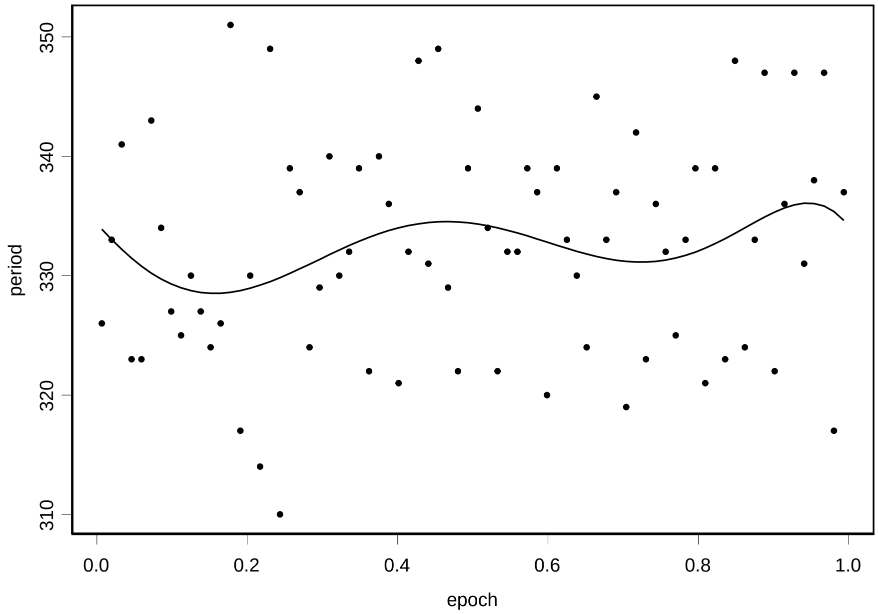

Figure 9. Data and polynomial fit for Mira

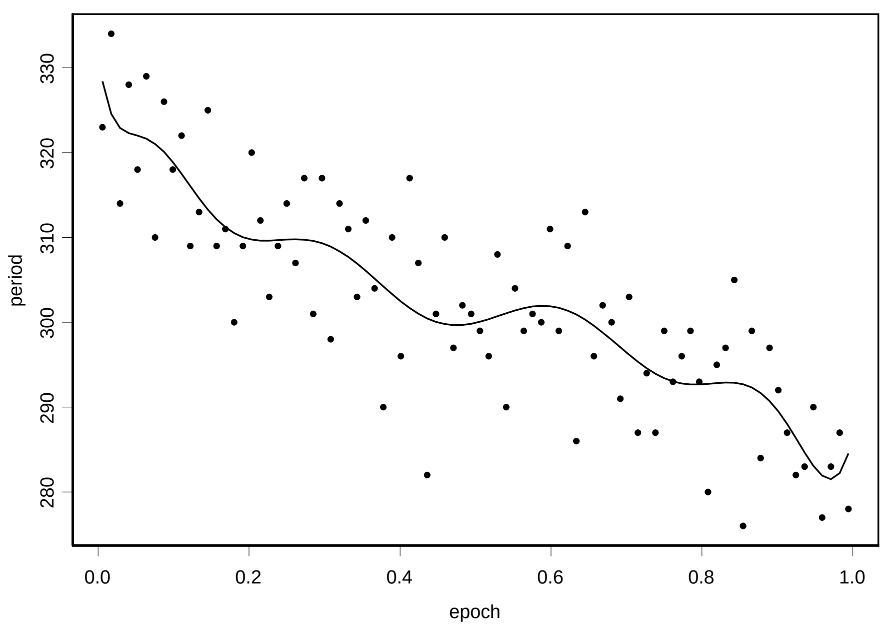

Figure 10. Data and polynomial fit for R. Aquilae

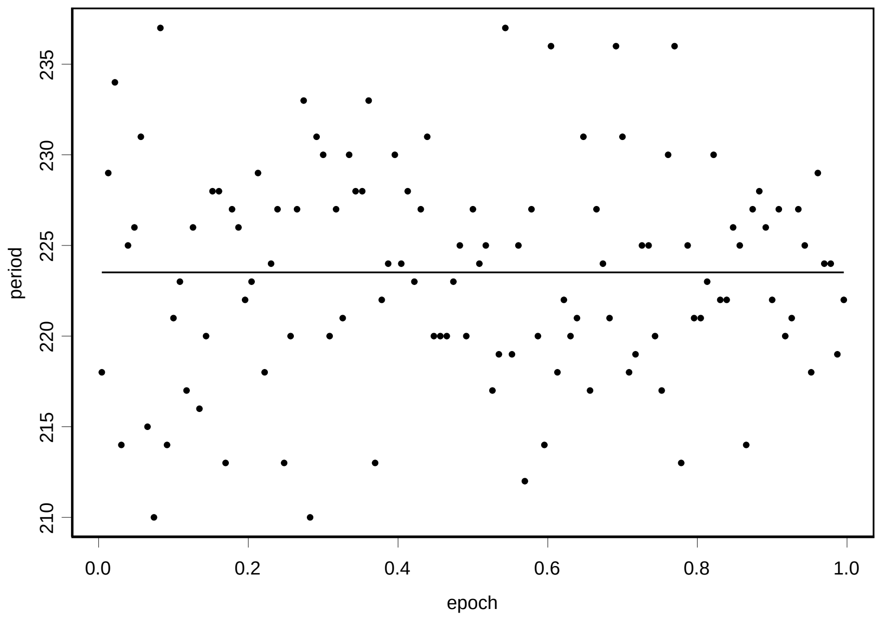

Figure 11. Data and polynomial fit for R. Bootis

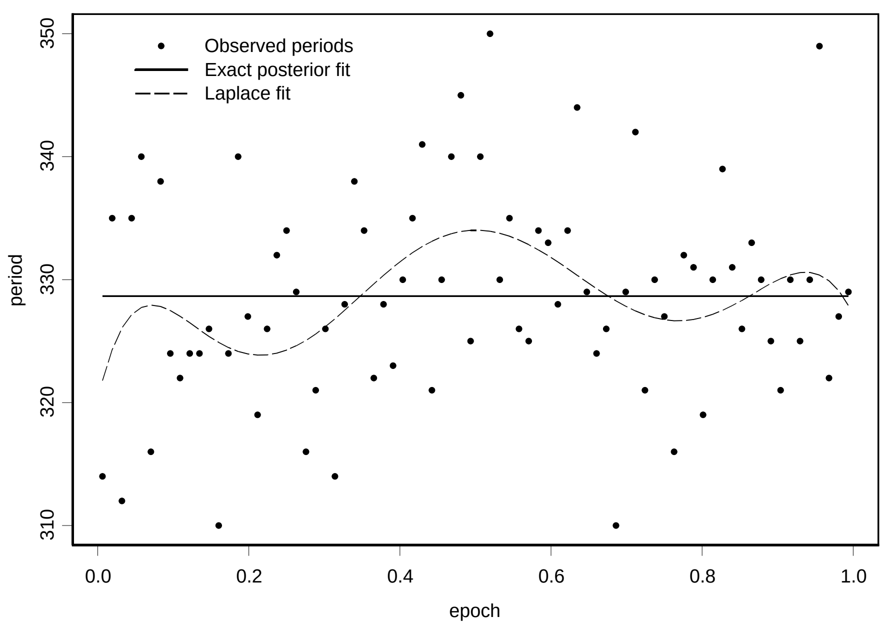

Figure 12. Data and polynomial fits for R. Canum Venaticorum

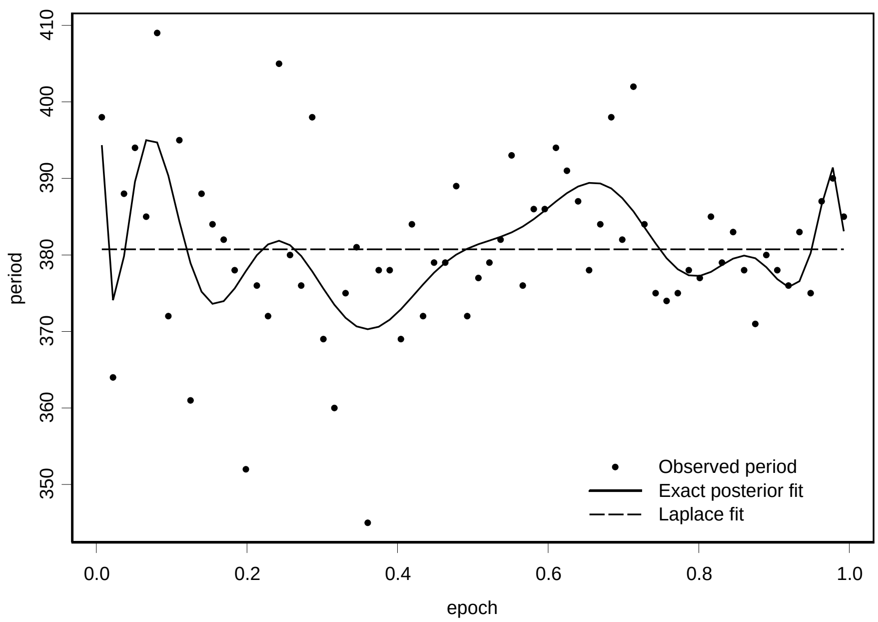

Figure 13. Data and polynomial fits for W. Aquarii

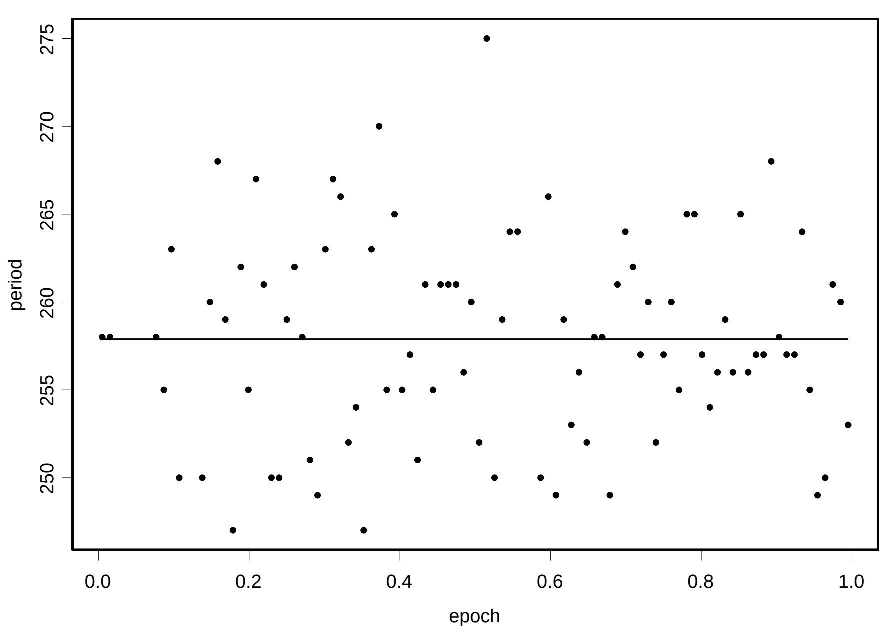

Figure 14. Data and polynomial fit for X. Draconis

5.2 Discussion

The variable stars analyzed, Mira ($n = 76$), R. Aquilae ($n = 86$), R. Bootis ($n = 115$), R. Canum Venaticorum ($n = 78$), W. Aquarii ($n = 68$) and X. Draconis ($n = 89$), represent a variety of different behaviors. For Mira, the model with highest posterior probability is $(m, h) = (6, 0)$ but models with nearby degrees are nearly as likely. For R. Aquilae, the model with highest posterior probability is $(m, h) = (11, 0)$ and no other model has significant posterior probability. For R. Bootis, R. Canum Venaticorum and X. Draconis the model with highest posterior probability is the model $(m, h) = (0, 1)$. However the details are different for these three stars. For R. Bootis, the $(m, h) = (0, 1)$ model has virtually all the posterior probability. For X. Draconis the posterior probability for $h = 1$ models falls off more slowly as the degree increases. For R. Canum Venaticorum, the posterior probability for $h = 1$ models has a secondary peak at degree $m = 4$ and is smaller but not insignificant for many other degrees. For W. Aquarii, the model with highest posterior probability is $(m, h) = (13, 1)$ but the degrees $m = 0$ and $m = 5$ also have appreciable posterior probability.

For these six variable stars, the Standard BIC generally selected the model with highest posterior probability, although BIC usually did not provide a good estimate for the posterior probability. The failure of the Standard BIC to provide a good estimate of the posterior probability was in agreement with our theoretical results as derived in Chapter IV.

Modified BIC also provided a very poor estimate of the posterior probabilities. This is not too surprising, since the constant terms that were derived in Chapter IV are not included in the Modified BIC. These constant terms are quite large. For example, in (4.28) the constant terms include

$$\prod_{i=1}^{m}(2i+1)^{1/2}\binom{2i}{i}$$

which grows roughly like $2^{m^2/2}$. Further, the Modified BIC has a much smaller penalty for $h = 1$ models with $\hat{\sigma}_Z^2 > 0.001$. As a result it is strongly biased towards the $h = 1$ models with low polynomial degree. Modified BIC chose the model with highest posterior probability only for stars R. Bootis, R. Canum Venaticorum and X. Draconis when the model with the highest posterior probability was the model with $(m, h) = (0, 1)$. In fact, with the exception of R. Aquilae, the Modified BIC always chose the $(m, h) = (0, 1)$ model.

The Standard Laplace, the Modified Laplace and the Asymptotic Laplace approximations generally gave comparable results for all six stars that were analyzed. All three Laplace approximations provided more accurate estimates of the posterior probabilities than their BIC counterparts. The asymptotic calculations used to derive the Asymptotic Laplace approximation all rely on the fact that the sample size n is much larger than the degree m of the polynomial in the model. This is reflected in the fact that the Asymptotic Laplace approximation provided poorer estimates for the posterior probabilities for high values of m.

An exception is the variable star R. Canum Venaticorum for which all three Laplace approximations did fairly poorly. The reason for this failure is that the Laplace approximations are based on expansions about the maximum likelihood estimators of the parameters but the main contribution to the exact posterior probability comes from near the posterior mode. For most stars the likelihood is sufficiently peaked and the prior density sufficiently flat that the MLEs and the posterior modes are close. However, R. Canum Venaticorum had maximum likelihood estimates $\hat{\sigma}_Z^2$

and $\hat{b}$ which were reasonably far from the mean values of these estimators over all the variable stars in the data set. Since our priors were based on the MLEs for the data, the prior is not flat near the MLE for R. Canum Venaticorum and the posterior modes were relatively far from the MLEs. This accounts for the failure of the Laplace approximations.

The Standard Laplace approximation did very poorly for $h = 1$ models with $\hat{\sigma}_Z^2$ near zero. This is not surprising, since we saw in Chapter IV that the Laplace approximation is not valid in this case. This weakness is not too serious since in this case the $h = 0$ model with the same polynomial degree has substantially higher posterior probability. As a result the exact posterior probabilities of $h = 1$ models with small $\hat{\sigma}_Z^2$ are small and therefore accurate estimates of them are not of great interest.

The Standard BIC and the Modified BIC both provided poor estimates of the posterior probabilities and hence their use in this manner is not recommended. However, the Standard BIC does seem to provide a fairly good criterion for model selection. This justifies our method of estimating priors, wherein we used parameter estimates corresponding to models that maximized BIC. The three Laplace-based approximations all performed satisfactorily. Since the Asymptotic Laplace approximation tended to deviate from the "exact" posterior probabilities for high degrees, it must be used with some care. All the estimates of the posterior probabilities required dramatically less computation time than the "exact" posterior probabilities. Computing "exact" posterior probabilities for all 378 variable stars in the data would be a prohibitively lengthy calculation.

CHAPTER VI

SIMULATIONS

The data on actual stars of Chapter V provide one verification of Laplace's method for approximating posterior probabilities. However, for actual data we do not know that the selected model is in fact correct. Therefore, we would also like to study the behavior of BIC and Laplace's method for data which are generated from the Star model described in Chapter III.

The first simulation was based on the fit to the variable star Mira with $(m, h) = (0, 1)$. The parameter values for θ_0 and η were the Mira MLEs for this model. The sample size was $n = 76$ as in the data on Mira. We generated 300 data sets with these parameters. For each replicate, we computed the BIC approximation to the posterior probability and the standard Laplace approximation to the posterior probability. Given the relatively large number of data sets, computing the exact posterior probabilities was not feasible. Further, the main interest is in seeing how well BIC and Laplace perform in choosing the correct degree, so it is not imperative to have the exact posteriors.

Recall that the exact posterior probabilities, BIC and the standard Laplace approximation for Mira all chose the model $(m, h) = (6, 0)$ as the best fit. As a result the parameters for the $(m, h) = (0, 1)$ fit have a relatively small $\sigma_Z^2 = 1.1125$ compared to a variance of roughly $\exp(2\beta_0) = 41.7$ for the MA part of the model. Thus this is a fairly challenging set of parameter values. This caused problems for both BIC and Laplace's method, since both often favored $h = 0$ models with higher degree over the correct model.

For this simulation, BIC did very poorly at predicting the true degree $m = 0$. In

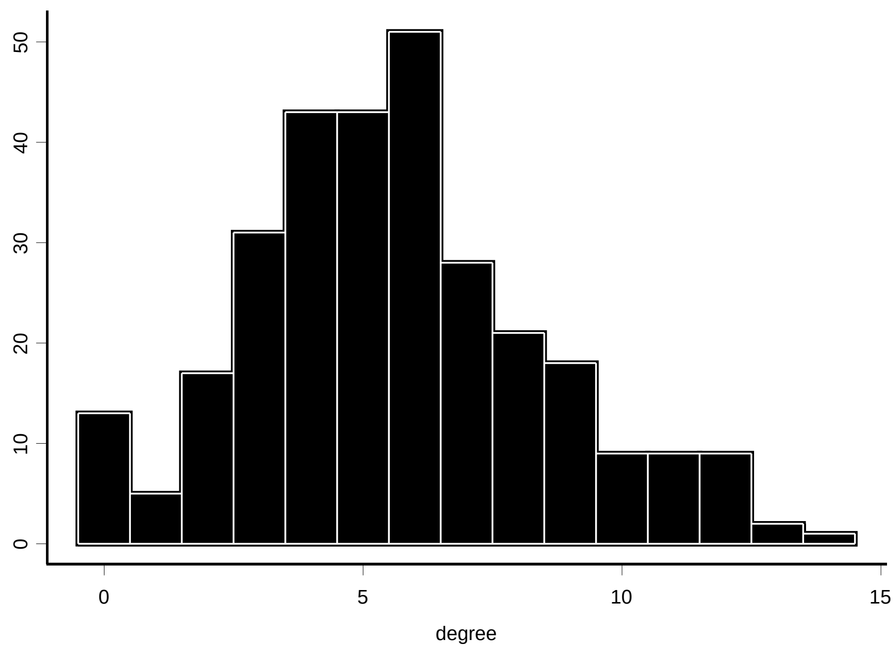

Figure 15. Histogram of degrees chosen by BIC for $n = 76$, $(m, h)_{true} = (0, 1)$

only 13 of the 300 replicates did BIC choose the correct degree. The correct degree was BIC's second choice in an additional 3 replicates. The average value of the BIC estimate of the posterior probability of degree 0 was only 0.0356. Only 7.67% of the replicates had estimated posterior probability greater than 0.10 and only 10.3% had estimated posterior probability greater than 0.05. This means that if we had used BIC to select the model for these simulated data sets, even with a fairly substantial bias in favor of the correct degree, we would have incorrectly concluded that a trend was present in the vast majority of the data sets. The actual degrees chosen are shown in Figure 15. Figure 16 shows a cumulative frequency plot of π_{BIC}, the BIC approximation to the posterior probability of no trend.

For this first simulation, the Laplace approximation did better than BIC, but still not satisfactorily at predicting the true degree $m = 0$. In only 132 of the 300 replicates

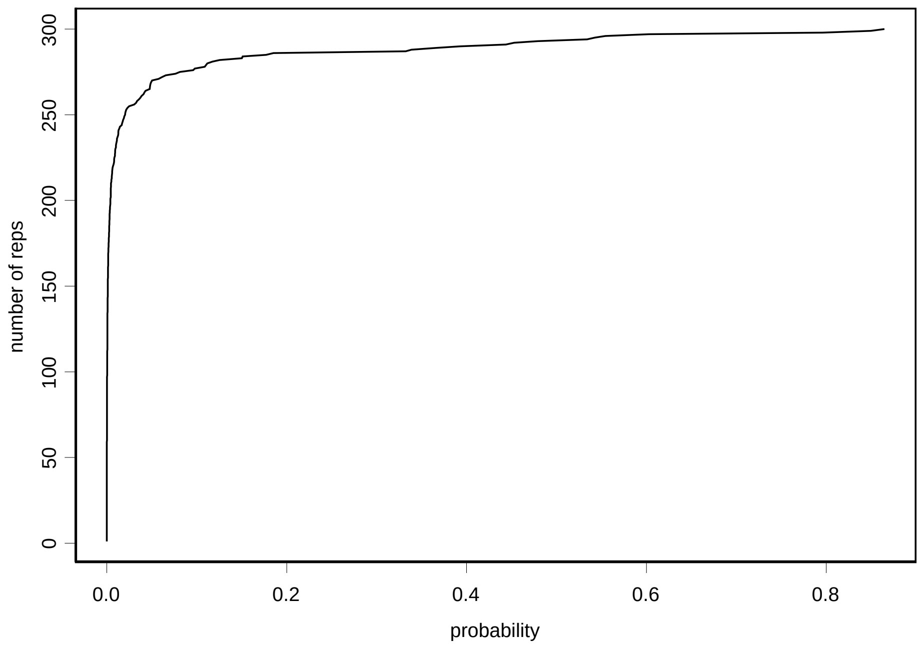

Figure 16. Number of replicates with $\pi_{BIC}(\text{deg. }0|\mathbf{Y}) < p$ *for* $n = 76$ *and* $(m,h)_{true} = (0,1)$

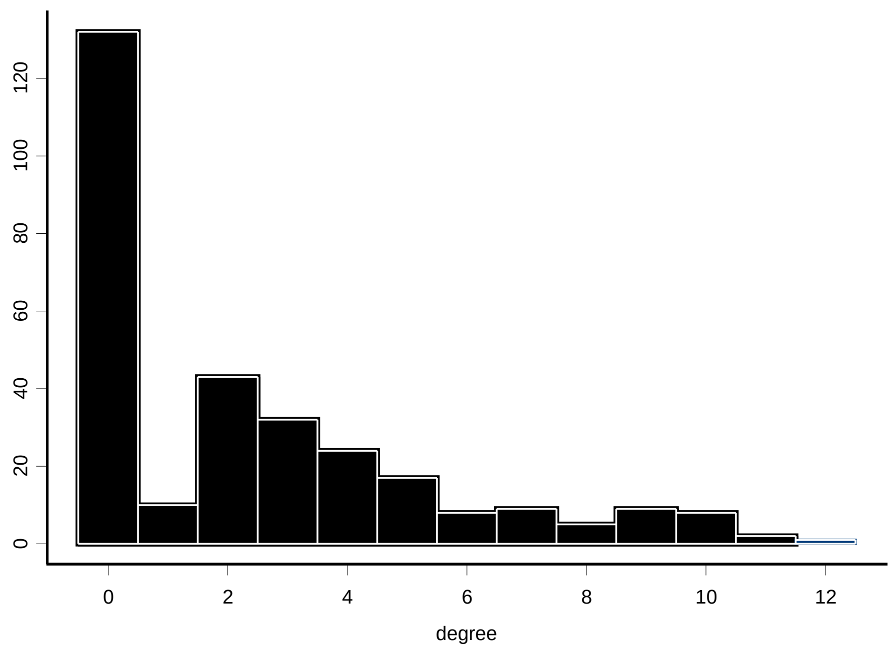

Figure 17. Histogram of degrees chosen by Laplace for $n = 76$ *and* $(m, h)_{true} = (0, 1)$

did Laplace choose the correct degree. The correct degree was Laplace's second choice in an additional 47 replicates. The average value of the Laplace approximation to the posterior probability of degree 0 was 0.3529. Only 65.7% of the replicates had approximate posterior probability greater than 0.10 and only 76.7% had approximate posterior probability greater than 0.05. Thus Laplace's method, with a strong bias in favor of no trend, would have correctly concluded no trend was present in the majority of the data sets, but still would have had a high error rate. The actual degrees chosen by the Laplace approximation are shown in Figure 17. Figure 18 shows a cumulative frequency plot of π_{Laplace}, the Laplace approximation to the posterior probability of no trend.

The sample size $n = 76$ used in the first round of simulations is not that large, especially since the models had as many as 20 parameters. Therefore a second batch

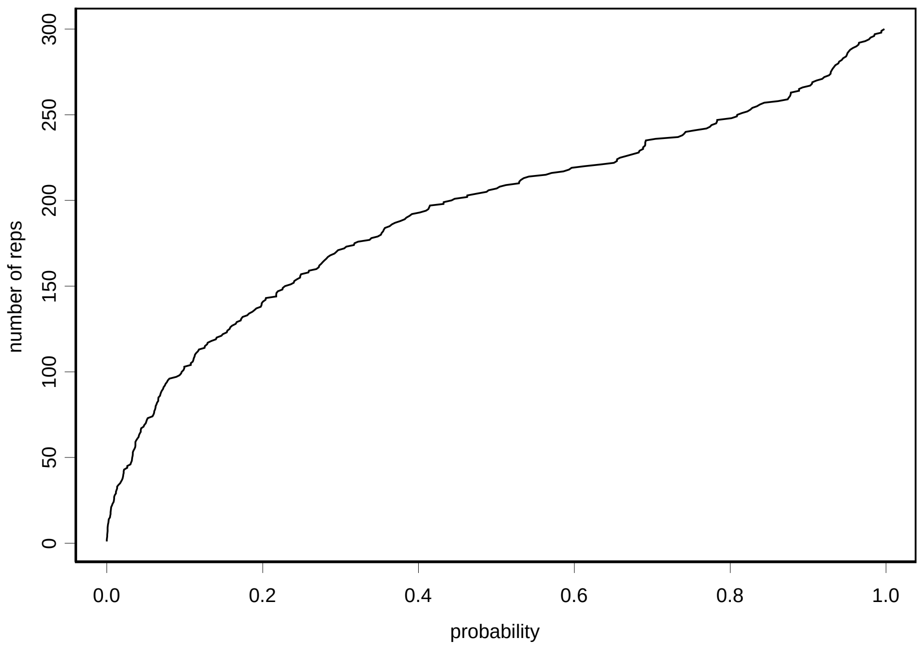

Figure 18. Number of replicates with $\pi_{Lap}(\text{deg. } 0|\mathbf{Y}) < p$ *for* $n = 76$ *and* $(m,h)_{true} = (0,1)$

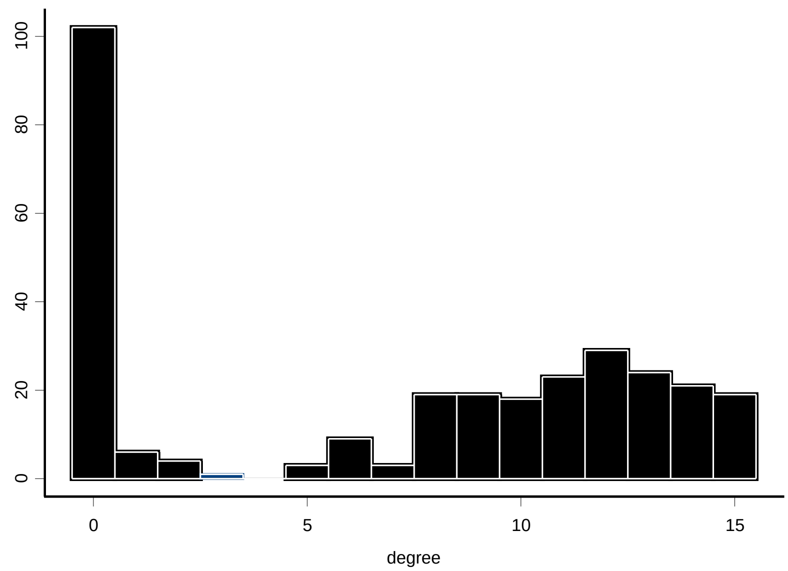

Figure 19. Histogram of degrees chosen by BIC for $n = 150$ *and* $(m, h)_{true} = (0, 1)$

of 300 data sets were generated with the same parameter values but with a sample size of $n = 150$. Both BIC and Laplace's method still chose $h = 0$ models with higher degree some of the time, but did dramatically better than for the smaller $n = 76$ data sets.

For these larger data sets, BIC chose the true degree $m = 0$ in 102 of the 300 replicates and the correct degree was BIC's second choice in an additional 32 replicates. The average value of the BIC estimate of the posterior probability of degree 0 was 0.2769. In addition, 46% of the replicates had estimated posterior probability greater than 0.10 and 50.7% had estimated posterior probability greater than 0.05. The actual degrees chosen are shown in Figure 19. Figure 20 shows a cumulative frequency plot of π_{BIC}, the BIC approximation to the posterior probability of no trend. While a substantial improvement over the results for BIC with $n = 76$, these results are still not quite as good as Laplace's method's results even for $n = 76$.

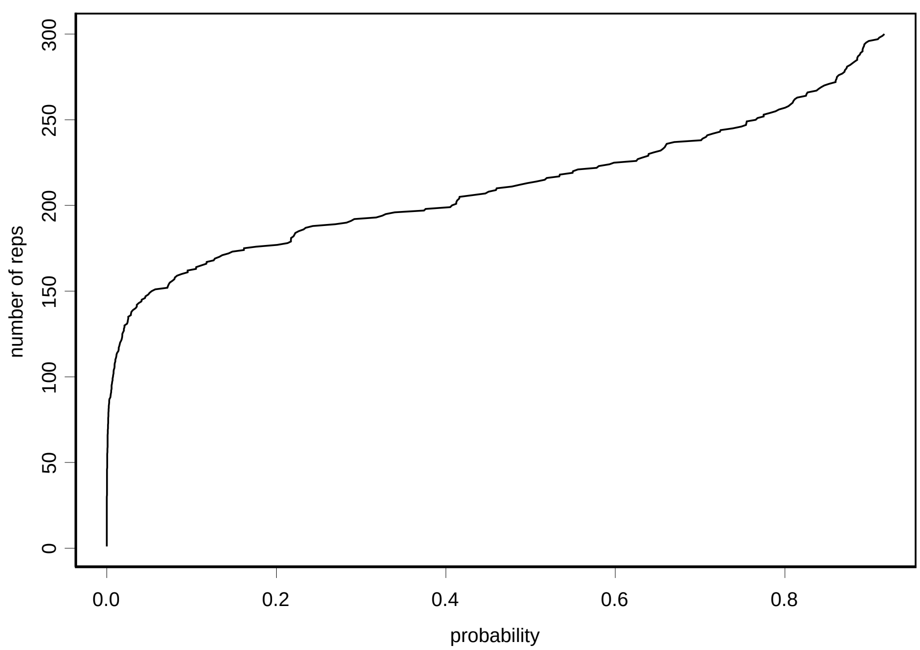

Figure 20. Number of replicates with $\pi_{BIC}(\text{deg. } 0|\mathbf{Y}) < p$ *for* $n = 150$ *and* $(m, h)_{true} = (0, 1)$

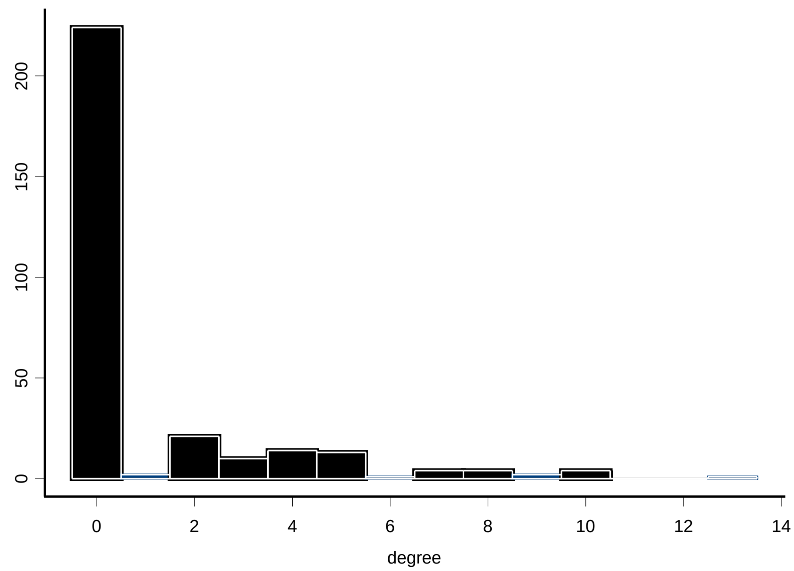

Figure 21. Histogram of degrees chosen by Laplace for $n = 150$ and $(m, h)_{true} = (0, 1)$

For these larger data sets, the Laplace approximation did very well. In 224 of the 300 replicates Laplace's method chose the correct degree. The correct degree was the second choice in an additional 40 replicates. The average value of the Laplace approximation to the posterior probability of degree 0 was 0.5805. For 89.7% of the replicates the approximate posterior probability was greater than 0.10 and for 92.3% the approximate posterior probability was greater than 0.05. The actual degrees chosen by the Laplace approximation are shown in Figure 21. Figure 22 shows a cumulative frequency plot of π_{Laplace}, the Laplace approximation to the posterior probability of no trend. Thus the performance of Laplace's method in this case is adequate. As expected, both Laplace's method and BIC improve greatly with larger sample sizes, but for intermediate sample sizes Laplace's method is substantially better than BIC.

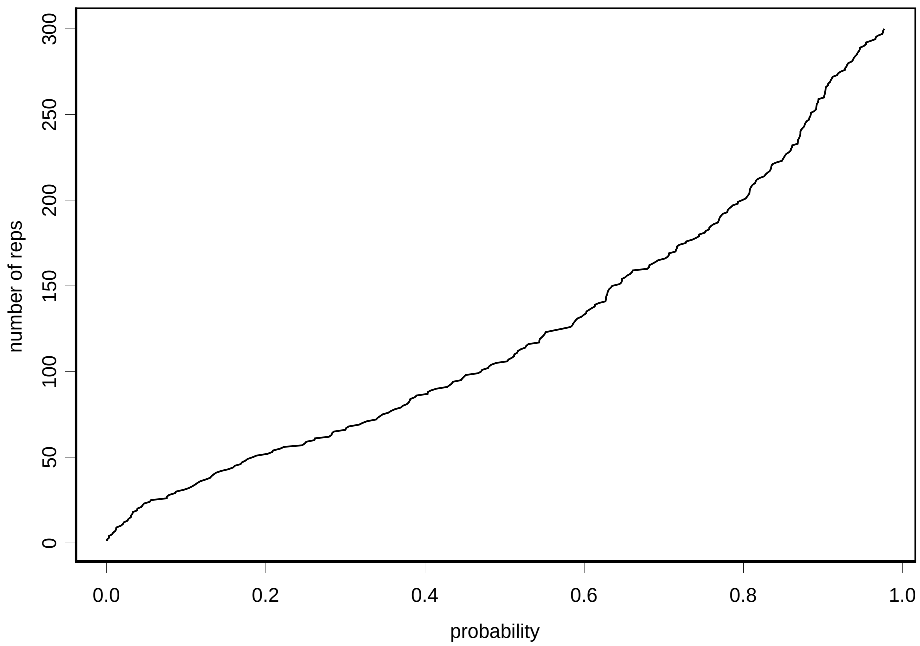

Figure 22. Number of replicates with $\pi_{Lap}(\text{deg. } 0|\mathbf{Y}) < p$ *for* $n = 150$ *and* $(m, h)_{true} = (0, 1)$

For the final simulation we chose a no trend model with $\sigma_I^2 = 0$, i.e., a $(m, h) = (0, 0)$ model. The MLEs for Mira for the $(m, h) = (0, 0)$ model are fairly far from the mean of $(\hat{\beta}_0, \hat{b})$. We saw from our analysis of R. Canum Venaticorum in Chapter V that when the MLEs are distant from the prior mode Laplace's method is much less accurate. Hence instead of choosing parameter values based on Mira, we chose the parameter values to be the prior mode. With these parameters values Laplace's method should provide a much better approximation to the exact posterior. The sample size was still taken to be $n = 76$. Again 300 data sets were generated.

For this simulation, both BIC and Laplace's method were nearly perfect. Since the true model had $\sigma_I^2 = 0$, neither method seemed inclined to choose models with $\sigma_I^2 > 0$. BIC chose the correct degree in 280 of the 300 replicates and never chose a degree larger than 2. The histogram of the degrees chosen is shown in Figure 23. Laplace's method chose the correct degree in all 300 replicates. The average value of the BIC estimate of the posterior probability of degree 0 was only 0.7721. The average Laplace approximation to the posterior probability of degree 0 was 0.9467. For BIC, 99.3% of the data sets had estimated posterior probability greater than 0.10. For Laplace's method all 300 replicates had approximate posterior probability greater than 0.10. Figures 24 and 25 show the cumulative frequency plots of π_{BIC} and π_{Laplace}.

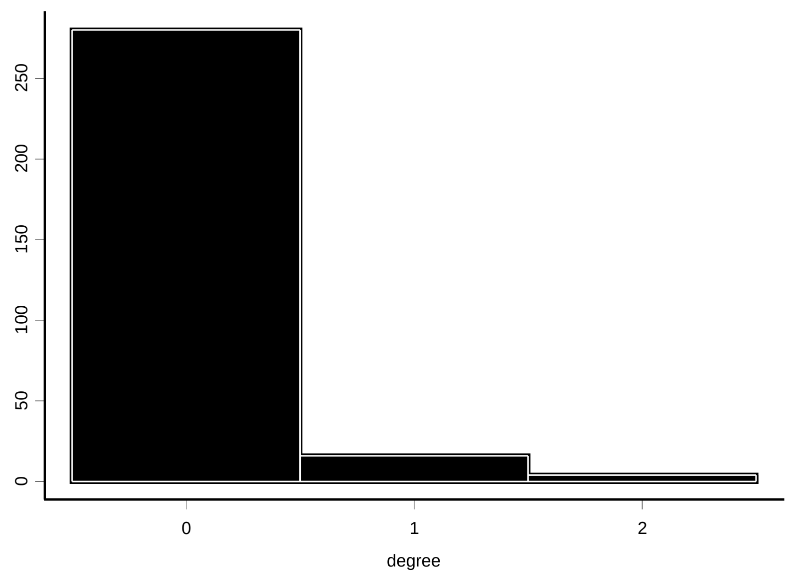

Figure 23. Histogram of degrees chosen by BIC for $n = 76$ *and* $(m, h)_{true} = (0, 0)$

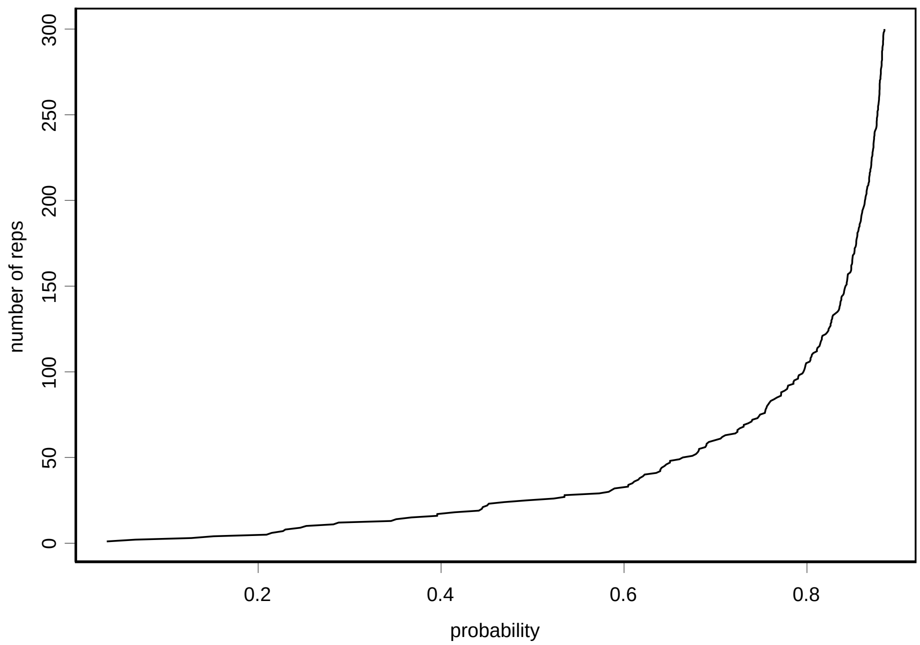

Figure 24. *Number of replicates with* $\pi_{BIC}(\text{deg. } 0|\mathbf{Y}) < p$ *for* $n = 76$ *and* $(m, h)_{true} = (0, 0)$

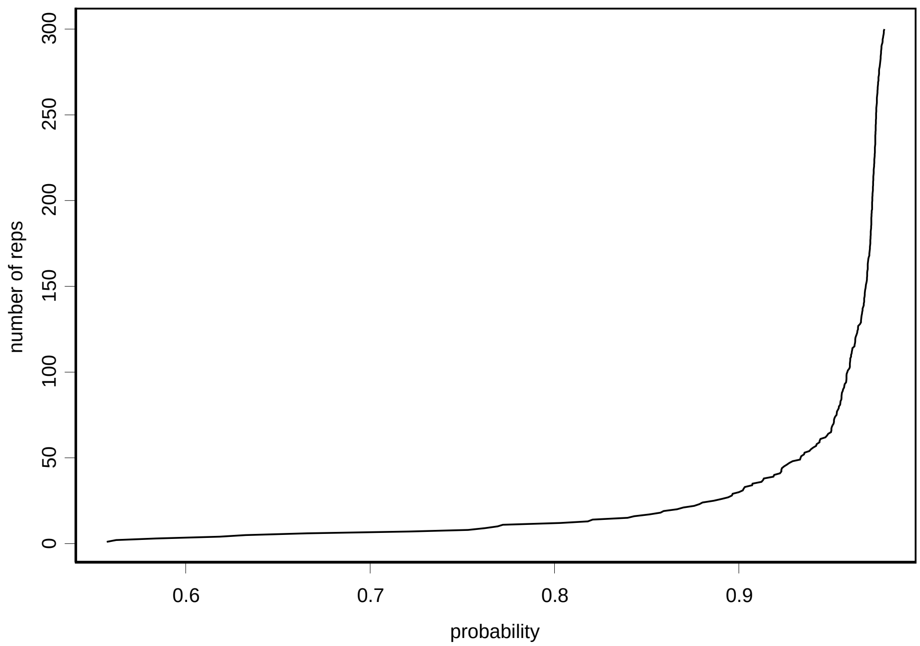

Figure 25. *Number of replicates with* $\pi_{Lap}(\text{deg. } 0|\mathbf{Y}) < p$ *for* $n = 76$ *and* $(m,h)_{true} = (0,0)$

REFERENCES

Akaike, H. (1973) "Information Theory and an Extension of the Maximum Likelihood Principle," in *Second International Symposium on Information Theory*. B. N. Petrov and B. F. Csaki, eds., Budapest: Academiai Kiado, 267-271.

Akaike, H. (1974) "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, 19, 716-723.

Akaike, H. (1976) "Canonical Correlation Analysis of Time Series and the Use of an Information Criterion," in *System Identification*. R. K. Mehra and D. G. Lainiotis, eds., New York: Academic Press, 27-96.

Akaike, H. (1977) "On Entropy Maximization Principle," in *Proceedings of the Symposium on Applications of Statistics*. P. R. Krishnaiah, ed., Amsterdam: North-Holland, 27-47.

Akaike, H. (1978) "On Newer Statistical Approaches to Parameter Estimation and Structure Determination," *International Federation of Automatic Control*, 3, 1877-1884.

Akaike, H. (1979) "A Bayesian Extension of the Minimum AIC Procedure of Autoregressive Model Fitting," *Biometrika*, 66, 237-242.

Bhansali, R. J. and Downham D. Y. (1977) "Some Properties of the Order of an Autoregressive Model Selected by a Generalization of Akaike's FPE Criterion," *Biometrika*, 64, 547-551.

Bozdogan, H. (1987) "Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytical Extensions," *Psychometrika*, 52, 345-370.

Cavanaugh, J. E. and Neath, A. A. (1999) "Generalizing the Derivation of the Schwarz Information Criterion," *Communications in Statistics - Theory and Methods*, 28, 49-66.

Hart, J. D., Koen, C., and Lombard, F. (2004) "An Analysis of Times between Local Maxima on the Light Curves of Variable Stars," under revision for *Annals of Statistics*.

Haughton, D. M. A. (1988) "On the Choice of a Model to Fit Data from an Exponential Family," *The Annals of Statistics*, 16, 342-355.

Johnson, R. A. (1967) "An Asymptotic Expansion for Posterior Distributions," *Annals of Mathematical Statistics*, 38, 1899-1906.

Johnson, R. A. (1970) "Asymptotic Expansions Associated with Posterior Distributions," *Annals of Mathematical Statistics*, 41, 851-864.

Kashyap, R. L. (1982) "Optimal Choice of AR and MA Parts in Autoregressive Moving Average Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4, 99-104.

Kass, R. E. and Raftery, A. E. (1995) "Bayes Factors," *JASA*, 90, 773-792.

Kass, R. E. , Tierney, L. and Kadane, J. B. (1988) "Asymptotics in Bayesian Computation," in *Bayesian Statistics 3*. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds., Oxford: University Press, 289-305.

Kass, R. E. , Tierney, L. and Kadane, J. B. (1989) "Approximate Methods for Assessing Influence and Sensitivity in Bayesian Analysis," *Biometrika*, 76, 663-674.

Kass, R. E. , Tierney, L. and Kadane, J. B. (1990) "The Validity of Posterior Expansions Based on Laplace's Method," in *Bayesian and Likelihood Methods in Statistics and Econometrics: Essays in Honor of George A. Barnard*. S. Geisser, J. S. Hodges, S. J. Press and A. Zellner, eds., Amsterdam: North-Holland, 473-488.

Kass, R. E. , Tierney, L. and Kadane, J. B. (1991) "Laplace's Method in Bayesian Analysis," in *Statistical Multiple Integration*. N. Flournoy and R. K. Tsutakawa, eds., Providence, RI: ASA, 89-99.

Kass, R. E. and Vaidyanathan, S. K. (1992) "Approximate Bayes Factors and Orthogonal Parameters with Application to Testing Equality of Two Binomial Proportions," *Journal of the Royal Statistical Society, Ser. B*, 54, 129-144.

Kass, R. E. and Wasserman, L. (1995) "A Reference Bayesian Test for Nested Hypotheses and Its Relationship to the Schwarz Criterion," *JASA*, 90, 928-934.

Laplace, P.S. De (1847) *Oeuvres*, 7, Paris: Imprimerie Royale.

Laplace, Le Marquis de (1820) *Théorie Analytique des Probabilités*, 3rd ed., Courcier, Paris. Reprinted in *Complete Works*, 7. Paris: Gauthier-Villars, 1886.

LeCam, L. (1953) "On Some Asymptotic Properties of Maximum Likelihood Estimates and Related Bayes' Estimates," *University of California Pub. Statist.*, 1, 277-329.

LeCam, L. (1956) "On the Asymptotic Theory of Estimation and Testing Hypothesis," in *Proceedings of the Third Berkeley Symposium*, 1. J. Neyman and E.L. Scott, eds., Berkeley: University of California Press, 129-156.

LeCam, L. (1958) "Les Propiétés Asymptotiques de Solutions de Bayes," *Pub. Inst. Statist. Univ. Paris*, 7, 17-35.

LeCam, L. (1966) "Likelihood Functions for Large Number of Independent Observations," in *Research Papers in Statistics. Festschrift for J. Neyman*. F. N. David, ed., New York: Wiley, 167-187.

Olver, F. (1997) *Asymptotics and Special Functions*, New York: Academic Press.

Pauler, D. K., Wakefield, J. C., and Kass, R. E. (1999) "Bayes Factors and Approximations for Variance Component Models," *JASA*, 94, 1242-1253.

Raftery, A. E. (1996) "Approximate Bayes Factors and Accounting for Model Uncertainty in Generalized Linear Models," *Biometrika*, 83, 251-266.

Schwarz, G. (1978) "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 461-464.

Tierney, L. and Kadane, J. B. (1986) "Accurate Approximations for Posterior Moments and Marginal Densities," *JASA*, 81, 82-86.

Tierney, L., Kass, R. E. and Kadane, J. B. (1987) "Interactive Bayesian Analysis Using Accurate Asymptotic Approximations," in *Computer Science and Statistics: 19th Symposium on the Interface*. R. Heideberger, ed., Alexandria, VA: ASA, 15-21.

Tierney, L., Kass, R. E. and Kadane, J. B. (1989a) "Fully Exponential Laplace Approximations to Expectations and Variances of Nonpositive Functions," *JASA*, 84, 710-716.

Tierney, L., Kass, R. E. and Kadane, J. B. (1989b) "Approximate Marginal Densities of Nonlinear Functions," *Biometrika*, 76, 425-433.

Wong, W. H. and Li., B. (1992) "Laplace Expansion for Posterior Densities of Nonlinear Functions of Parameters," *Biometrika*, 79, 393-398.

VITA

Suriani Pokta

Department of Statistics

447 Blocker Building

Texas A&M University

College Station, TX 77843-3143

Education

Ph.D. in statistics, Texas A&M University, College Station, 2004.

M.A. in statistics, University of California, Berkeley, 1994.

B.S. in mathematics/applied science, University of California, Los Angeles, 1992.

Professional Experience

1996-1997 Biostatistician II, Applied Logic Associates, Inc., TX.

1995-1996 Biostatistician I, Quintiles, Inc., Mountain View, CA.

1994-1995 Statistical Programmer, Quintiles, Inc., Mountain View, CA.

1993-1994 Consulting Statistician, Chiron Corporation, Emeryville, CA.

1992 Graduate Student Instructor, Department of Statistics - UC Berkeley, CA.

Honors

Sherwood Prize for Outstanding Senior in Mathematics

Magna Cum Laude

Dean's Honors List

Provost's Honors List